Option Pricing in Random Field Models with Stochastic Volatility for the Term Structure of Interest Rates

Baowei Xu

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operation Research.

Chapel Hill
2011

Approved by:

Chuanshu Ji, Advisor

Amarjit Budhiraja, Committee Member

Jan Hannig, Committee Member

Eric Renault, Committee Member

Shankar Bhamidi, Committee Member

Abstract

BAOWEI XU: Option Pricing in Random Field Models with Stochastic Volatility for the Term Structure of Interest Rates.
(Under the direction of Chuanshu Ji.)

In this dissertation, we introduce a general interest rate modeling framework by looking at yield curves in a Hilbert space, and bridge the popular HJM factor models with more recent random field models. Then we study the problem of vanilla interest rate option (cap) pricing under the random field model. This will be a generalization of Kennedy (1994) paper in the sense that the volatility will also follow a random field process instead of being deterministic. In particular, we consider both cases in which the two random fields for forward rates and volatilities are independent or correlated. In the computation of option prices, we have proposed a log-normal approximation of the summary statistics - integrated volatility, for the independent case and have proposed a trivariate Gaussian approximation for the correlated case. The approximations will enable us to compute option prices much faster than the usual brute force Monte Carlo method which introduces certain discretization error. Finally, we perform simulation studies of a MCMC estimation procedure for a special random field model with one-factor stochastic volatility.

Acknowledgments

Special thanks to the distinguished faculty members who served on my committee: Professors Chuanshu Ji, Amarjit Budhiraja (chair), Jan Hannig, Eric Renault, Anh Le and Shankar Bhamidi. As my advisor, Dr. Ji provided detailed guidance and encouragement throughout the course of preparing for and conducting the research. Especially thank Dr. Ji for the numerous weekends he spent over the phone advising me during the past year when I was away from school. Dr. Le served faithfully on the committee until travel prevented him from attending the defense. I am grateful for the helpful comments he provided on the draft. Dr. Bhamidi kindly filled in for Dr. Le, and he provided insightful comments on short notice. Thanks to all my committee members for their support, patience, encouragement, and useful suggestions. Thanks to Prof. Jingfang Huang for suggestions on high dimensional computation. And thanks to the attendees of the 2009 JSM program and the UNC STAT-OR internal seminars for their useful comments. Thanks to the whole statistics department for all the course work and support I have received over the years, without which nothing would have been possible.

Thanks to my mother, sister and others in the big family for their belief and pride in this accomplishment. Thanks to many classmates and friends who cheered me on from the beginning, especially Xingye Qiao, Feng Liu, Ruiwen Zhang, Xin Liu, Jun Ge and Wenjie Chen. And lastly, special thanks to my wife Zhixing Zhang for the continuous support and encouragement to finish this work.

Table of Contents

Chapter 1

Introduction

Modeling the term structure dynamics of (stochastic) interest rates is crucial in the research of fixed-income markets. However, problems in this area are much more complicated than their counterparts in equity (stock) markets. There have been several stages in the development of interest rate models. Early studies were focused on the dynamics of short rate $r(t)$ at time t, such as Vasicek (1977), Cox, Ingersoll and Ross (1985), Hull and White (1993). Then term structure models came along, particularly the HJM framework established in the seminal work of Heath, Jarrow and Morton (1992), which governs the dynamics of forward rate $f(t, T)$ set at time t while becoming effective at $T > t$, or equivalently of bond price $P(t, T)$ or yield $Y(t, T)$. Such an extension is interpreted as a transition from finite state models to infinite state models, because at each t, $r(t)$ is a 1D quantity while $f(t, T)$ is a function (a yield curve) over different terms $T \in [t, t + T_{max}]$. Here T_{max} is some fixed long term say 30 years. A short rate model yields a term structure as an output, which may inevitably be different from observed yield curve data. Such discrepancy can be easily avoided in a term structure model, which reads the original term structure $\{f(0, T)\}$ as input data. Term structure models have their own limitations. For one thing, assuming the evolution of $f(t, T)$ follows a stochastic differential equation (SDE) driven by a 1D standard Wiener process (Brownian motion), then the increments $df(t, T_1)$ and $df(t, T_2)$ with the same t but different terms T_1 and T_2 will be perfected correlated. This problem can be alleviated by adding more "factors", i.e. one can assume in a HJM model that $df(t, T)$ is driven by a k-dimensional Wiener process with $k > 1$. The idea is that although there are infinitely many possible values for the term T, one hopes the information about a yield curve can be summarized by reducing to just a finite number of determining factors. Such dimension reduction is undoubtedly useful, but the choice of k is an art: too large the model will be very difficult to calibrate, while if too small it won't have enough flexibility to fully represent market data.

A finite factor HJM model allows us to fit the current yield curve, but it does not permit consistency with term structure innovation. Therefore, practitioners need to continuously recalibrate model parameters in order to fit new term structures. Since those parameters are not supposed to be stochastic processes and constantly updated, the HJM framework cannot resolve its inconsistency with empirical data. A more recent approach, called random field models or string models, makes recalibration of parameters unnecessary as it can be consistent with any observed term structure, and it enjoys several other modeling advantages as well. Meanwhile, it invites a daunting challenge in required computation and empirical studies.

The introduction of random field models represents another qualitative change, from finite factor models to infinite factor models. This is a relatively new area. Kennedy (1994), (1997) originally proposed Gaussian random fields for modeling forward rates and presented an explicit option pricing formula. Goldstein (2000), Santa-Clara and Sornette (2001) contributed to set a more general framework. The encyclopedic book on interest rate modeling, James and Webber (2000), included a section on random field models. Mathematically, a random field model is driven by an infinite dimensional Wiener process (or equivalently its "derivative" referred to as a white noise process).

For every t, the cumulative random noise $W(t,\cdot) = \{W(t,T) : \; t \le T \le t + T_{max}\}$ is defined as a Gaussian random variable taking values in a Hilbert space H with a prescribed correlation function. The correlation function of $W(t,\cdot)$, along with a further specification of drift and volatility processes will determine the dynamics of $f(t,T)$ via a SDE. In particular, $df(t,T)$ and $df(t',T')$ in this model are correlated in a non-degenerate and non-perfect manner for any non-identical pairs (t,T) and (t',T'), which makes it more flexible than the previous factor models. See the book Carmona and Tehranchi (2006) for the most updated rigorous treatment of random field interest rate models. It should be mentioned that the infinite dimensional white noise setting enables us to view HJM factor models as a special case: using the orthogonal series expansion of $W(t,\cdot)$ in H, we can truncate it to a finite sum with k terms that naturally give rise to a k-factor HJM model.

We propose to study random field term structure models in this dissertation. Here are several possible novel contributions we wish to make.

(i) *Volatility modeling*

Volatility modeling is always at the center of financial econometrics, and its importance is recognized by researchers and practitioners in the community of fixed-income markets. We adopt stochastic volatility (SV) models in which the volatility $\{V(t,T)\}$ is also defined as a random field in addition to the forward rate field $\{f(t,T)\}$. SV models are used extensively in economics and finance. Early works include Clark (1973), Taylor (1982), Tauchen and Pitt (1983), and Gallant, Hsieh, and Tauchen (1991). A review was provided by Ghysels, Harvey, and Renault (1996), and a lot more applications have appeared since then. However, there is virtually none related to random field interest rate models, except for Collin-Dufresne and Goldstein (2003), to the best of our knowledge. Even in Collin-Dufresne and Goldstein (2003), only a finite factor affine SV model was adopted so that option pricing can be done by using the Fourier inversion formula proposed in Heston (1993). In contrast, we treat $\{V(t,T)\}$ as a genuine random field in our work.

(ii) *Option pricing and probability approximation schemes*

Our primary interest is to derive option pricing formulas for random field SV term structure models. Although random field models enjoy an advantage of modeling flexibility, very little has been done for providing workable option pricing formulas, due to the tremendous technical challenge. Kennedy (1994) gave an explicit formula in the case of Gaussian fields for forward rate with deterministic volatility. Collin-Dufresne and Goldstein (2003) introduced a finite factor affine SV model and offered the resulting Heston's formula for option pricing. We propose a bold approach here to define both forward rate $\{f(t,T)\}$ and SV $\{V(t,T)\}$ as random fields. Moreover, we adopt a log-linear SV structure instead of the square-root affine SV structure. Therefore, Heston's formula is not applicable to the setting we consider. Using log-linear SV models has a long history. Its applications were showed, for example, in Scott (1987), Danielson (1994), Jacquier et al. (1994), Kim et al. (1998), Chernov et al. (2003), Cheng et al. (2008). A probability approximation scheme was proposed in Cheng et al. (2008) and its promise was demonstrated in both simulation studies and an empirical example with foreign exchange return and option data. The complexity in random field models makes it more imperative to develop similar approximation methods. We tackle this problem in two steps. Step one concerns a special case without a "leverage effect", i.e. we assume the two white noise processes for forward rates and SV are uncorrelated. The term "leverage effect" is borrowed from the equity markets where a negative price shock will increase the leverage of a firm and thus more likely a higher volatility. In other words, there exists a negative correlation between the innovation for equity price and its volatility, while economical interpretation is more difficult in the fixed income field. This simplification allows us to extend Kennedy's pricing formula by using a log-normal random variable as a proxy for the integrated volatility (denoted as σ_t^2) in the life time $[t, T]$ of an option. This is a significant dimension reduction in Monte Carlo integration: brute-force simulation of random field samples is replaced by generating a 1D quantity σ_t^2. Specifications of the two parameters in the proposed log-normal density are transformed from the first and second moments of σ_t^2, which are calculated with a couple of low dimensional integrals involved. See Chapter 3 for details. While some authors argue that innovations in interest rate levels are largely uncorrelated with innovations in the volatility of interest rates (e.g., Ball and Torous (1999), Chen and Scott (2001), and Heidari and Wu (2003)), Trolle and Schwartz(2009) argues that the correlation between forward rates and volatility innovations is important in capturing the implied volatility skewness, which is an import feature that has been observed in the market. In Step two, we will study this more realistic but sophisticated case with a "leverage", i.e. the aforementioned two random fields of white noise are assumed to be correlated. Technically, we encounter a serious challenge in the derivation of an approximate pricing formula. Instead of a single summary statistic σ_t^2, there are seven variables involved in the conditional expectation expression for the option (cap) price, which fortunately can be further grouped into three summary statistics. Still, we follow the same path by computing all first and second moments, based on which we will specify a trivariate Gaussian law for the summary random vector. The option price can be computed by simulating a large number of Gaussian vectors which approximate the distribution of the summary random vectors. We use numerical study to validate the proposed (log) Gaussian approximation scheme, which turns out to be very good under both cases.

(iii) *Model calibration simulation study*

Numerical studies are further behind the theoretical development in random field term structure models. Besides simulation studies for computing an option price, a major statistical task is model calibration using, ideally, a combined data structure of asset returns and option prices. Here returns in the fixed income field are based on yields or zero-coupon bond prices, which are both equivalent to forward rates $\{f(t,T)\}$. Examples of bond options include caps and swaptions. Again, there have been a lot of previous works in statistical inference for SV models with stock returns, we only mention Chernov et al. (2003) and Chib et al. (2002) here. They represent the GMM/EMM frequentist approach and the Markov Chain Monte Carlo (MCMC) Bayesian approach respectively. With the combined data of stock returns and options, see Pan (2002) for the use of GMM, Eraker (2004) and Cheng et al. (2008) for MCMC. As for inference on random field models considered in this work, we are only aware of Bester (2004) as a rare attempt at performing MCMC Bayesian computational inference using interest rate data. In the mean time, there are no published papers yet on random field model-based inference using both yields and option data. To reduce the complexity with MCMC computational inference, we will focus on simulation study of calibrating a one-factor SV random field model using forward rates data. And defer the model calibration for a random field volatility model using both forward rates and option data into future studies. In regard to the specification of the random field, Longstaff et al. (2001) suggested some ad hoc specifications for correlation functions. Although they did not assume SV, the suggestions may shed some light on how we should pursue our inference.

The rest of the dissertation is organized as follows: A rigorous mathematical framework for random field term structure models is introduced in Chapter 2. Chapter 3 concerns our main contributions, in which we derive option pricing formulas and their numerical computation through probabilistic approximation. Also included are numerical studies that examine the accuracy and speed of the approximation schemes. Chapter 4 contains an extension to Chapter 3, where we consider correlated instead of independent innovations between forward rates and their volatilities. Finally in Chapter 5 we carry out a simulation study to test the performance of a MCMC calibration procedure for a random field forward rate process with one factor SV. Here a calibration procedure has been repeatedly performed on simulated forward rates data and parameters estimation performance has been examined.

Chapter 2

Term Structure Dynamics

Modeling the dynamics of term structure is essential for pricing interest rate derivatives such as cap (European call option on short rate) and swaption (option on coupon bond). There are currently two main approaches in the literature: one is the traditional finite factor models of HJM type [Heath et al (1992)], driven by a finite number of Brownian motions; the other is the more recent development of random field models, also called string models [Kennedy (1994,1997), Goldstein (2000), Santa-Clara and Sornette (2001), Collin-Dufresne and Goldstein (2003)] String models have infinite factors, hence enjoy greater modeling flexibility.

In this chapter, we present a general framework for string models where random noises are defined as an infinite dimensional Wiener processes. Here is the motivation: A forward rate or bond price is indexed by (t, T), where t represents the time at which these financial variables are recorded, and T stands for the maturity date. Random noises introduced at different pairs (t, T) will be correlated, modeled by a certain correlation function. We will define a Wiener process $\mathbf{W} = \{\mathbf{W}(t), t \geq 0\}$ such that at each t, the random variable $\mathbf{W}(t) = \{\mathbf{W}(t, t+u) := \mathbf{W}(t)(u),\ 0 \leq u \leq T_{max}\}$ takes values in a Hilbert space H consisting of square integrable functions on $[0, T_{max}]$, $L^2(0, T_{max})$, and is governed by a Gaussian measure over H. The covariance operator tQ for the Gaussian measure will characterize the dependence among noises of different terms T. Having defined this string of noises, a stochastic drift term and volatility coefficient will be introduced so that the forward rate process will follow a stochastic differential equation evolving in time t. Ultimately, dependence among forward rates at different pairs (t, T) will depend on both the covariance operator Q and the proposed volatility dynamics V.

Note that a new contribution in this work is to assume the stochastic volatility $V = \{V(t, T)\}$ also follows a random field model, and to propose an approximation scheme to price vanilla options written on zero-bond prices or interest rates. The reason why we need a stochastic volatility model beyond ARCH/GARCH in term structure is because several recent studies have shown that there seem to be risk factors that affect the prices of caps and swaptions but not the underlying Libor and swap rates. In other words, bonds do not seem to span interest rate derivatives. e.g. Heidari and Wu (2003), Collin-Dufresne and Goldstein (2003), Li and Zhao(2006). Being popular SV models, there is a serious issue with factor models: high dimensional factor models with SV are very difficult to calibrate, and thus it is rare to see models with more than three-factors. However, Jagannathan, Kaplin and Sun (2003) suggest that low-dimensional affine models can not capture the joint dynamics of yield, caps and swaptions; Dai and Singleton (2003) also find that observed innovation correlation between non-overlapping intervals of yield curve is much lower than a two or three factor model would suggest. Random field model is naturally high dimensional, and in fact it is regarded as infinite factor model. With some structural assumptions, it may be easier to implement than a factor model. Longstaff, Santa-Clara and Schwartz (2001) started with a random field representation and applied Principle Component Analysis to match parameters with the first M-components from data. However, Kerkhof and Pelsser(2002) pointed out that random field model estimated using this manner are observationally equivalent to the finite-factor models.

This chapter is organized as follows: in the first section, we introduce some probability basics and define function valued Wiener process by introducing general Gaussian measure; then we bridge the popular factor HJM model with the more recent random field model under a general framework, and provide conditions under which they can both be special cases of the general framework; in the third section, we will examine certain special specifications of the random field model so that it can degenerate to some popular models in the literature; and lastly, to complete the model specification, we will summarize some of the common specifications for volatility.

2.1 Probability Basics

Here we follow J. Zabczyk's notes on 'Parabolic Equations on Hilbert Spaces' to introduce some probability theory and then define the Gaussian measure and Wiener process in a general Hilbert space.

Probability Space

A measurable space $(E, \mathcal{E})$ consists of a set E and of a σ-field $\mathcal{E}$. If μ is a nonnegative measure on a measurable space $(E, \mathcal{E})$ such that $\mu(E) = 1$ then μ is called a probability measure, and the triplet $(E, \mathcal{E}, \mu)$ is called a probability space. If $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ are two measurable spaces, then any measurable transformation $X : \Omega \mapsto E$ is called a random variable. Assume that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and X is a random variable taking values in E. The image μ of the measure $\mathbf{P}$ by the transformation X, $\mu(A) = \mathbf{P}(\omega : X(\omega) \in A), \forall\ A \in \mathcal{E}$ is called the law or the distribution of X and denoted by $\mathcal{L}(X)$.

Assume that H is a separable Hilbert space equipped with the inner product $< \cdot, \cdot >$ and Borel σ-field $\mathcal{B}(H)$. Probability measures on H will always be regarded as being defined on $\mathcal{B}(H)$. If μ is a probability measure on H then its characteristic function G_μ is a complex valued function on H of the form

$$G_\mu(\lambda) = \int_H e^{i<\lambda,x>} \mu(dx), \quad \lambda \in H.$$

There exists a one to one correspondence between characteristic functions and the probability measures, which means $G_\mu(\lambda) = G_\nu(\lambda), \forall \lambda \in H$ if and only if the two measures μ and ν are identical.

Gaussian Measures

Gaussian probability measure on R is well known and usually is defined by its density or characteristic function. More generally a measure μ on a Hilbert space H is Gaussian if all linear mappings $y \mapsto < \lambda, y >$ defined on H, considered as random variables on $(H, \mathcal{B}(H), \mu)$ with values in $(R, \mathcal{B}(R))$, have Gaussian laws. Random variables are Gaussian if their laws are Gaussian. Similar to the Gaussian measure on R, we can also define the measure on H through its characteristic function with the following theorem:

Theorem 1. *A measure μ is Gaussian if and only if*

$$G_\mu(\lambda) = e^{i<\lambda,m> - \frac{1}{2}<Q\lambda,\lambda>}, \quad \lambda \in H$$

where $m \in H$ and Q is a self-adjoint, nonnegative, operator with finite trace. We will denote the measure as $N(m, Q)$ with mean m and covariance Q.

Assume that X is a Gaussian random variable $\mathcal{L}(X) = N(m, Q)$, and $\{e_k\}$ is the complete system of eigenfunctions of Q corresponding to eigenvalues $\{\gamma_k\}$, which means $Qe_k = \gamma_k e_k, \quad k = 1, 2, \dots, \infty$. Then we have the following facts:

Fact 1: $E < X, \lambda >=< m, \lambda >, \quad \forall \lambda \in H$;

Fact 2: $E|X - m|^2 = \sum_{k=1}^{\infty} \gamma_k$;

Fact 3: $E < X - m, a >< X - m, b >=< Qa, b >, \quad \forall a, b \in H$;

Fact 4: if $\gamma_k > 0$ then $w_k = \gamma_k^{-\frac{1}{2}} < X, e_k >$ is a standard normal variable with real values, $\mathcal{L}(w_k) = N(0, 1)$, and the random variables $w_k, k = 1, \dots, \infty$ are mutually independent.

Fact 5: $X = \sum_{k=1}^{\infty} \sqrt{\gamma_k} w_k e_k + m$ in distribution.

The requirement that Q has finite trace will make sure that the infinite series in the Fact 5 convergences a.e. in H. This can be seen from the following corollary.

Corollary 1. *let $\{e_k\}$ be an orthonormal sequence in a Hilbert space H, $\{w_k\}$ be a sequence of independent Gaussian random variables with $\mathcal{L}(w_k) = N(0, 1), \forall k = 1, 2, \dots, \infty$, and let $\gamma_k, k = 1, \dots, \infty$ be nonnegative numbers. Then the series*

$$\sum_{k=1}^{\infty} \sqrt{\gamma_k} w_k e_k$$

converges in $H, \mathcal{P} - a.s.$ if and only if $\sum_{k=1}^{\infty} \gamma_k < \infty$.

It can be seen that, Fact 1 relates to the mean of a Gaussian measure; Fact 2 relates to the overall variance of the measure; and Fact 3 actually characterizes the covariance structure of the Gaussian measure, which can be seen more clearly under the following condition and corollary:

Condition 1: We will consider a special linear operator Q which is self-adjoint, non-negative, with finite trace, and also has kernel function $c(\cdot,\cdot)$ such that $\forall\ a \in H,\ Qa \in H$ and $(Qa)(\cdot) = \int c(s,\cdot)a(s)ds$. We will call $c(\cdot,\cdot)$ the kernel of the operator Q.

Corollary 2. *Suppose $X = \{X(s), 0 \le s \le T_{max}\} \in H$ is a Gaussian measure with $\mathcal{L}(X) = N(m,Q)$. If Q satisfies* ***Condition 1*** *with kernel function $c(\cdot,\cdot)$ then $Cov(X(s),X(t)) = c(s,t)$.*

Proof: Without loss of generality we can assume $m = 0$. For any $x_1, x_2 > 0$ and $\Delta > 0$ if take $a = I_{[x_1,x_1+\Delta]}(s)$ and $b = I_{[x_2,x_2+\Delta]}(t)$, then from Fact 3 we have

$$E < X, a >< X, b > \; = \; < Qa, b >$$

$$\Updownarrow$$

$$E \int_{x_1}^{x_1+\Delta} X(s)ds \int_{x_2}^{x_2+\Delta} X(t)dt \; = \; \int_{x_2}^{x_2+\Delta} Qa(t)dt$$

Now since $Qa(x) = \int c(x,y)a(y)dy$, we have :

$$\begin{aligned}\int_{x_1}^{x_1+\Delta}\int_{x_2}^{x_2+\Delta} Cov(X(s),X(t))dsdt &= \int_{x_2}^{x_2+\Delta}\int c(s,t)a(s)dsdt \\ &= \int_{x_2}^{x_2+\Delta}\int_{x_1}^{x_1+\Delta} c(s,t)dsdt\end{aligned}$$

Because x_1, x_2 and Δ are arbitrary, we obtain $Cov(X(s),X(t)) = c(s,t)$. $\qquad QED.$

Infinite Dimensional Wiener Process

Here and for the rest of the dissertation, we will assume the string shock to be an infinite dimensional Wiener process $\mathbf{W} = \{\mathbf{W}(t);\ t \geq 0\}$, taking values in a separable Hilbert space $H = C[0, T_{max}])$ equipped with inner product $< a, b >:= \int_0^{T_{max}} a(s)b(s)ds, \forall a, b \in H$. Here it is essentially assumed that forward rate curve and its instantaneous shock are both continuous functions of time to maturity, which is a reasonable assumption.

Suppose $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space with a given increasing family of σ-fields $\mathcal{F}_t \subset \mathcal{F}, t \geq 0$. A family $\mathbf{W}(t), t \geq 0$ of H-valued random variables is called a Wiener process, if and only if,

1. $\mathbf{W}(0) = 0$;
2. For almost all $\omega \in \Omega, \mathbf{W}(t, \omega)$ is a continuous function of t;
3. $\mathbf{W}(t_1), \mathbf{W}(t_2) - \mathbf{W}(t_1), \ldots, \mathbf{W}(t_n) - \mathbf{W}(t_{n-1})$ are independent random variables, $\forall 0 \leq t_1 < t_2 < \ldots < t_n, \forall n \in N$
4. $\mathcal{L}(\mathbf{W}(t) - \mathbf{W}(s)) = \mathcal{L}(\mathbf{W}(t-s)), \forall t \geq s$.

If $\mathbf{W}$ is a Wiener process, then $\forall t > 0$, $\mathcal{L}(\mathbf{W}(t)))$ is a Gaussian measure on H with mean 0 and covariance tQ, where $Q \in \mathcal{H}$ is a non-negative operator satisfies ***Condition 1***. Suppose Q has eigenvalues $\{\gamma_k\}(\forall k, \gamma_k$ is non-negative and real number) with corresponding eigenfunctions $\{e_k\}$. Then the characteristic function for $\mathbf{W}(t)$ will be

$$G_{\mathbf{W}(t)}(\lambda) = Ee^{i<\mathbf{W}(t), \lambda>} = e^{-\frac{1}{2}t<Q\lambda, \lambda>}, \lambda \in H.$$

Moreover, utilizing the definition of Wiener process and the Fact 1-5 of Gaussian measure, the following facts corresponding to Wiener process $\mathbf{W}(t)$ can be shown easily:

Fact 1: $E < \mathbf{W}(t), a >= 0, \quad \forall a \in H;$

Fact 2: $E < \mathbf{W}(t), \mathbf{W}(s) >= E|\mathbf{W}(t \wedge s)|^2 = (t \wedge s) \sum_{k=1}^{\infty} \gamma_k;$

Fact 3: $E\left[< \mathbf{W}(t), a >< \mathbf{W}(s), b >\right] = (t \wedge s) < Qa, b >, \quad \forall a, b \in H;$

Fact 4: If $\gamma_k > 0$ then $w_k(t) = \gamma_k^{-\frac{1}{2}} < \mathbf{W}(t), e_k >$ is a one-dimensional standard Wiener process, and the Wiener processes $w_k(t), k = 1, \ldots, \infty$ are mutually independent.

Fact 5: $\mathbf{W}(t) = \sum_{k=1}^{\infty} \sqrt{\gamma_k} w_k(t) e_k \quad \forall t \geq 0$ in distribution.

Finite Rank Approximation

Corollary 3. *Define*

$$\mathbf{W}^N(t) = \sum_{k=1}^{N} \sqrt{\gamma_k} w_k(t) e_k, \forall t \geq 0, \forall N = 1, \ldots, \infty.$$

Then for $\forall$ arbitrary $T > 0$ there exists a sequence $N_m \to \infty$ such that $\mathbf{W}^{N_m}(.)$ is uniformly convergent on $[0, T]$ as $m \to \infty$, for almost everywhere $\omega \in \Omega$.

Proof: Let $N > M$ then

$$E\left(\sup_{t \leq T} |\sum_{k=M+1}^{N} \sqrt{\gamma_k} w_k(t) e_k|^2\right) = E\left(\sup_{t \leq T} \sum_{k=M+1}^{N} \gamma_k w_k^2(t)\right)$$
$$\leq \sum_{k=M+1}^{N} \gamma_k E(\sup_{t \in [0,T]} w_k^2(t)) \leq C \sum_{k=M+1}^{N} \gamma_k$$

where $C = E(\sup_{t \in [0,T]} w_1^2(t))$. Therefore, one can find an increasing sequence $\{N_m\}$

s.t.

$$P\left(\sup_{t\le T}|\mathbf{W}^{N_{m+1}}(t)-\mathbf{W}^{N_m}(t)|\ge\frac{1}{2^m}\right)\le\frac{1}{2^m},\quad m=1,\ldots,\infty$$

The result now follows. $\quad QED.$

Equivalent conditions under which general Wiener process has finite rank N, i.e. $\mathbf{W}(t)=\sum_{k=1}^{N}\sqrt{\gamma_k}w_k(t)e_k$:

1. $Qe_k=\gamma_k e_k$, and $\gamma_k=0,\ \forall k\ge N$;

2. $c(u,v)=Cov(\mathbf{W}(1)(u),\mathbf{W}(1)(v))=\sum_{k=1}^{N}\gamma_k e_k(s)e_k(t)$, this means the kernel of Q can be factorized; It can be seen if $c(t,t)=1,\ \forall 0\le t\le T_{max}$ then $\sum_{k=1}^{N}\gamma_k e_k^2(t)=1,\ \forall 0\le t\le T_{max}$. This additional constrain will decrease the number of tuning functions for $c(\cdot,\cdot)$, e_k, by one.

2.2 General HJM Models

$(\Omega,\mathcal{F},\mathcal{P})$ is a probability space with a given increasing family of σ-fields $\mathcal{F}_t\subset\mathcal{F},t\ge 0$. If we denote the whole forward rate curve with time to maturity between 0 and T_{max} at time t as $\mathbf{f}(t)$, then we can define the risk-neutral forward rate dynamics as:

$$\mathbf{f}(t)=\mathbf{f}(0)+\int_0^t\mu(s)ds+\int_0^t\sigma(s)d\mathbf{W}(s)\tag{2.1}$$

or equivalently in differentiate form,

$$d\mathbf{f}(s)=\mu(s)ds+\sigma(s)d\mathbf{W}(s)\tag{2.2}$$

where

- $\forall s,\ \mu(s)$ and $\mathbf{f}(s)$ both take value in a separable Hilbert Space H $(C[0,T_{max}])$, and $\mu(\cdot)$ is integrable a.e., $\mathrm{P}(\int_0^t |\mu(s)|ds < \infty) = 1, \forall t > 0$

- $\forall s, \sigma(s) \in \mathcal{H} := \mathcal{L}(H, H)$ is a linear operator from H to H which satisfies $P(\int_0^t ||\sigma(s)||^2_{\mathcal{H}} ds < \infty, \forall t \geq 0) = 1$. $\mathcal{L}(H, H)$ is the set of all linear operators from H to H, which from theory of functional analysis, is also a separable Hilbert space. This condition will ensure that $\int_0^t \sigma(s)dW(s)$ is well defined. Note that, most of the time in specific examples, we will further assume $\sigma(s)$ to satisfy ***Condition 2*** below.

- $\mathbf{W}(s)$ is an infinite dimensional Wiener process taking values in H with covariance $Q \in \mathcal{H}$,where Q is an operator satisfying ***Condition 1***.

- $\forall s \geq 0$, $\mu(s)$ and $\sigma(s)$ are adapted to the filtration of $\mathcal{F}_s$;

Condition 2[1]***:*** $\forall s \geq 0$, $\sigma(s)$ is a linear operator from H to H, such that $\forall a \in H$, $\sigma(s)a \in H$ and $(\sigma(s)a)(\cdot) = \sigma(s, \cdot)a(\cdot)$.

No-arbitrage Drift Condition

To exclude arbitrage opportunity, we should impose the following relationship between its drift and volatility under the risk-neutral measure:

Corollary 4. *Under* ***Condition 1*** *and* ***Condition 2****, the drift term of the forward rate dynamics in (2.2) should be identified from the volatility and correlation kernel of the Wiener process by the following equation:*

$$\mu(s, T) = \sigma(s, T) \int_s^T \sigma(s, u)c(T, u)du \tag{2.3}$$

---

[1]This condition has restrict the possible forward rate models by a great amount. It will make the forward rate of any time to maturity, including short rate, to be driven by a single noise source, thus can only produce a subset of finite factor HJM models as its special cases.

Proof: Here we briefly rewrite the Goldstein(2000) result. By definition, zero-coupon bond with maturity T will have price at time t:

$$P(t,T) = e^{-\int_t^T f(t,u)du}.$$

Differentiating with respect to time t, we obtain:

$$\frac{dP(t,T)}{P(t,T)} = r(t)dt - \int_t^T df(t,u)du + \frac{1}{2}\left[\int_t^T df(t,u)du\right]^2 \tag{2.4}$$

where $r(t) = f(t,t)$ is the short rate at time t. From the fundamental theorem of asset pricing, discounted asset price should be a martingale under risk-neutral measure. Apply this to the discounted bond price we obtain:

$$\frac{d\left\{e^{-\int_0^t r(u)du}P(t,T)\right\}}{e^{-\int_0^t r(u)du}P(t,T)} = \frac{dP(t,T)}{P(t,T)} - r(t)dt$$

from which we can see that since discounted bond price is a martingale, the bond price itself $\frac{dP(t,T)}{P(t,T)}$ should have a drift of short rate $r(t)$. Apply this into the dt term of (2.4), we get

$$\int_t^T \mu(t,u)du = \frac{1}{2}\left[\int_t^T df(t,u)du\right]^2$$

Taking derivative with respect to term T, we can have

$$\begin{aligned}
\mu(t,T) &= \left[\int_t^T df(t,u)du\right] df(t,T) \\
&= \int_t^T \sigma(t,u)d\mathbf{W}(t,u)\sigma(t,T)d\mathbf{W}(t,T)du \\
&= \sigma(t,T)\int_t^T \sigma(t,u)c(u,T)du
\end{aligned}$$

QED.

2.2.1 Finite-factor (rank) Approximation

From the finite rank approximation for infinite dimensional Wiener process section, we have that: $\forall \epsilon > 0$ there exists a positive constant K such that $E(\sup_{0\le s\le T}|\mathbf{W}(s) - \mathbf{W}^K(s)|^2) < \epsilon, a.e.$ Then equation (2.2) can be approximated by:

$$
\begin{aligned}
d\mathbf{f}(s) &= \mu(s)ds + \sigma(s)d\mathbf{W}(s) \\
&= \mu(s)ds + \sigma(s)\sum_{k=1}^{\infty}\sqrt{\gamma_k}e_k dw_k(s) \\
&\approx \mu(s)ds + \sum_{k=1}^{K}\sigma_k^*(s)dw_k(s) \\
&:= \mu(s)ds + \sum_{k=1}^{K}\sigma_k^*(s)dw_k(s) \\
&:= d\mathbf{f}^K(s)
\end{aligned}
\tag{2.5}
$$

where $\sigma_k^*(s) = \sqrt{\gamma_k}\sigma(s)e_k$ belongs to the Hilbert space H. Clearly $\mathbf{f}^K$ follows the K-factor HJM model, which is an approximation to the general infinite model. By specifying the variance operator $\sigma(s)e_k = \gamma_k^{-\frac{1}{2}}\sigma_k^*(s),\ \forall k \le K$ and $\sigma(s)e_k = 0,\ \forall k > K$, the general model will be the K-factor HJM model exactly.

If we look at the expected squared error of this approximation,

$$
\begin{aligned}
E|\mathbf{f}(t) - \mathbf{f}^K(t)|^2 &= E|\sum_{k=K+1}^{\infty} \sqrt{\gamma_k} \int_0^t \sigma(s) e_k dw_k(s)|^2 \\
&\leq \sum_{k=K+1}^{\infty} \gamma_k \int_0^t |\sigma(s) e_k|^2 ds \\
&\leq \sum_{k=K+1}^{\infty} \gamma_k \int_0^t ||\sigma(s)||_{\mathcal{H}}^2 ds
\end{aligned} \tag{2.6}
$$

$\forall$ fixed t, since $\int_0^t ||\sigma(s)^2||_{\mathcal{H}} ds$ is finite and $\sum_{k=1}^{\infty} \gamma_k \leq \infty$, then $E|\mathbf{f}(t) - \mathbf{f}^K(t)|^2 \to 0$, as $K \to \infty$. Or, we can also say that, $\forall$ fixed T, $\forall \epsilon$, there $\exists$ K, s.t. $\sup_{t \leq T} E|\mathbf{f}(t) - \mathbf{f}^K(t)|^2 < \epsilon$.

This property that infinite dimensional Wiener process can be approximated by a finite serial of one dimensional Wiener processes has laid ground for the success of finite factor HJM models. This is equivalent to say the covariance of the Wiener process can be approximated by a finite-ranked one. However, note that this is different from the Longstaff, Santa-Clara and Schwartz (2001) and Han(2007), where instead of finite factor approximation for instantaneous correlation matrix, they have used principal component analysis technique in the covariance matrix of forward rate in their implementation of the random field model in discrete time. In other words, they used finite factors to approximate $\int_t^{t+\Delta} \sigma(s) d\mathbf{W}(s)$ instead of $\mathbf{W}(t+\Delta) - \mathbf{W}(t)$ as in the current framework.

Comparing the two approaches, finite factors approximation for the standard Wiener process will have the property that instantaneous correlation and variance functions can be estimated separately. This is advantageous, since the market has a good feeling about how instantaneous shock correlations and how variances across different maturity look like. Also the estimated parameters are more interpretable. This is compatible with LIBOR market model, where correlation and variance are sometimes specified and estimated separately. Of course, since covariance is what can be observed directly, approximation aiming at it can usually have better fit in historical data, however, it gives less insight into possible future movements and is less interpretable to users.

2.2.2 Full-rank Specification

Replacing $\mathbf{W}$ by a finite dimension $\mathbf{W}^K$ is equivalent to say we would like to replace its covariance Q by a lower ranked Q^K. However, besides principle component analysis to approximate Q, which is finite ranked in nature, we can also approximate the correlation kernel of the Wiener process, $c(\cdot,\cdot)$, by certain parametric functions. Instead of having finite rank, the new correlation function will have full rank but controlled by only finite parameters. This alternative has the advantage of being infinite factor, i.e. no finite assets can complete the market even under deterministic volatility, but can still be estimated. The disadvantage is that we have to resort to the more complex two-dimensional stochastic analysis, which is not widely used in the finance literature. The third section of this chapter will introduce some of the proposed full-rank functional form for the correlation.

Special Note on Notations

In the following of this paper, we will denote finite dimension vector of Wiener process as $W(t)$, general Wiener process in the Hilbert space H as $\mathbf{W}(t)$, and one dimensional Wiener Process as $w(t)$. Depending on the situation for the Wiener Process, $\sigma(t)$ will be a vector of same length as $W(t)$, it can also be a member of $\mathcal{L}(H,H)$ for the H-valued $\mathbf{W}(t)$.

2.3 Some Specific Models

2.3.1 Heath-Jarrow-Morton (HJM) Factor Model

Suppose $W(t)$ is a n-dimensional Wiener process, otherwise under similar conditions of last section, we have the following HJM forward rate model:

$$f(t,T) = f(0,T) + \int_0^t \mu(s,T)ds + \int_0^t \sigma(s,T)^T dW(s)$$

or equivalently in differentiate form,

$$df(s,T) = \mu(s,T)ds + \sigma(s,T)^T dW(s).$$

The short rate process under the model will be:

$$r_t = f(t,t) = f(0,t) + \int_0^t \mu(s,t)ds + \int_0^t \sigma(s,t)^T dW(s)$$

and the differentiate form will be:

$$dr_t = \mu(t,t)dt + \sigma(t,t)^T dW(t) + \frac{\partial f(t,T)}{\partial T}|_{T\to t}.$$

We can also derive the dynamics for the zero bond prices from $B_t(T) = e^{-\int_t^T f(t,s)ds}$ and Ito's lemma:

$$
\begin{aligned}
d\int_t^T f(t,s)ds &= -f(t,t)dt + \int_t^T df(t,s)ds \\
&= -r_t dt + \int_t^T \mu(t,s)dt + \sigma(t,s)^T dW(t)ds \\
&= -r_t dt + \int_t^T \mu(t,s)dsdt + \int_t^T \sigma(t,s)^T dsdW(t) \\
&:= -r_t dt + \mu^*(t,T)dt + \sigma^*(t,T)^T dW(t)
\end{aligned}
$$

$$\Longrightarrow$$

$$
\begin{aligned}
dB_t(T) &= B_t(T)\{-d\int_t^T f_t(s)ds + \frac{1}{2}d^2\int_t^T f_t(s)ds\} \\
&= B_t(T)\{r_t dt - \mu^*(t,T)dt - \sigma^*(t,T)^T dW(t) + \frac{1}{2}\sigma^*(t,T)^T\sigma^*(t,T)dt\}
\end{aligned}
$$

which is equivalent to $\frac{dB_t(T)}{B_t(T)} = (r_t - \mu^*(t,T) + \frac{1}{2}\sigma^*(t,T)^T\sigma^*(t,T))dt - \sigma^*(t,T)^T dW(t)$

No-arbitrage Condition

Under the equivalent martingale measure, the discounted bond price should be a martingale. Since $dD(t) = -r_t D(t)dt$, then:

$$
\begin{aligned}
\frac{d(D(t)B_t(T))}{D(t)B_t(T)} &= -r_t dt + (r_t - \mu^*(t,T) + \frac{1}{2}\sigma^*(t,T)^T\sigma^*(t,T))dt - \sigma^*(t,T)^T dW(t) \\
&= (-\mu^*(t,T) + \frac{1}{2}\sigma^*(t,T)^T\sigma^*(t,T))dt - \sigma^*(t,T)^T dW(t) \qquad (2.7)
\end{aligned}
$$

Now from no-arbitrage, $D(t)B_t(T)$ need to be a martingale under the risk-neutral measure, which means that the drift term in the above SDE should be 0. That will give the drift condition for the factor HJM under risk-neutral measure:

$$\mu^*(t,T) = \frac{1}{2}\sigma^*(t,T)^T\sigma^*(t,T) \qquad (2.8)$$

We can also write the model in a random field framework as the following:

$$\begin{aligned} df(s,T) &= \mu(s,T)ds + \sigma^T(s,T)dW(s) \\ &= \mu(s,T)ds + \sqrt{\sigma^T(s,T)\sigma(s,T)}d\mathbf{W}(s,T) \end{aligned} \tag{2.9}$$

where $d\mathbf{W}(s,T) = \frac{\sigma^T(s,T)dW(s)}{\sqrt{\sigma^T(s,T)\sigma(s,T)}}$ can be seen as a string shock with correlation structure

$$\begin{aligned} d\mathbf{W}(t,T_1)\ d\mathbf{W}(t,T_2) &= \frac{\sigma^T(s,T_1)\sigma(s,T_2)}{\sqrt{\sigma^T(s,T_1)\sigma(s,T_1)}\ \sqrt{\sigma^T(s,T_2)\sigma(s,T_2)}}\ dt \\ &:= c(t,T_1,T_2)\ dt \end{aligned}$$

Note that in this case, the correlation function of the shock between different terms T_1 and T_2 at time t may depend on time t except the case when $\sigma_k(t,T) = \sigma_0(t,T)\ e_k(T),\ \forall k = 1,\ldots,K$. This is because *Condition 2* has been imposed implicitly here.

2.3.2 Hull-White Model

With the assumption of short rate process under equivalent measure,

$$dr_t = (\theta(t) - a(t)r_t)dt + \sigma(t)dW(t) \tag{2.10}$$

Hull and White(1990) showed that the price for a zero-coupon bond with maturity time T will be:

$$P_t(T) = A(t,T)e^{-B(t,T)r_t} \tag{2.11}$$

The coefficients satisfy

$$B(t,T) = \frac{B(0,T) - B(0,t)}{B_t(0,t)}$$
$$lnA(t,T) = ln\frac{A(0,T)}{A(0,t)} - B(t,T)\frac{A_t(0,t)}{A(0,t)} - \frac{1}{2}B^2(t,T)B_t^2(0,t)\int_0^t \frac{\sigma^2(s)}{B_s^2(0,s)}ds$$

with initial values for $A(0,T), B(0,T)$ determined by initial spot rate curve $r_0(T)$ and initial spot rate volatility $\sigma_0(T)$,

$$B(0,T) = r_0(T)\frac{\sigma_0(T)}{\sigma(0)}T$$
$$A(0,T) = P_0(T)e^{B(0,T)r_0}$$

Forward rate can be derived from bond prices. Pang(2000) has shown that:

$$\begin{aligned} cov(f(t_1,T_1), f(t_2,T_2)) &= B_{T_1}(0,T_1)B_{T_2}(0,T_2)\int_0^{t_1\wedge t_2} \frac{\sigma^2(s)}{B_s^2(0,s)}ds \\ &\equiv g(T_1,T_2)h(t_1 \wedge t_2) \qquad (2.12) \end{aligned}$$

2.3.3 Some Special String Shocks

An important element of random field model is the construction of string shocks. Under our framework of last section, it is equivalent to say we would like to specify the correlation structure of the Wiener process. To construct the String shock, we can assume the string to be a stochastic process along the T direction, or we can assume the string has to satisfy some SPDE.

Stochastic String follows O-U process

Suppose the string U(t) follows an Ornstein-Uhlenbeck(O-U) process

$$dU(t) = -\kappa U(t)dt + dB(t) \tag{2.13}$$

where κ is a positive constant. Then the string has a solution that $U(t) = \int_0^t e^{-\kappa(t-s)}dB(s)$. The covariance structure for this string will be like:

$$cov(U(t), U(s)) = \frac{1}{2\kappa}(e^{-\kappa|s-t|} - e^{-\kappa(s+t)}) \tag{2.14}$$

and the correlation function will take the form: $c(u, v) = e^{-\kappa(u-v)}\sqrt{\frac{1-e^{-2\kappa v}}{1-e^{-2\kappa u}}}$ when $u \geq v$.

Stochastic Strings as Solutions of SPDEs

If we assume that the stochastic string can be characterized by a linear second order SPDE:

$$\begin{aligned} &A(t,x)\frac{\partial^2 X(t,x)}{\partial t^2} + 2B(t,x)\frac{\partial^2 X(t,x)}{\partial t \partial x} + C(t,x)\frac{\partial^2 X(t,x)}{\partial x^2} \\ &\quad = H(t,x,X(t,x), \frac{\partial X(t,x)}{\partial t}, \frac{\partial X(t,x)}{\partial x}) \end{aligned}$$

where we restrict our discussion to linear equations

$$\begin{aligned} &H(t,x,X(t,x), \frac{\partial X(t,x)}{\partial t}, \frac{\partial X(t,x)}{\partial x}) \\ &\quad = D(t,x)\frac{\partial X(t,x)}{\partial t} + E(t,x)\frac{\partial X(t,x)}{\partial x} + F(t,x)X(t,x) + S(t,x) \end{aligned}$$

where $S(t,x)$ is some random "source" term. Santa-Clara and Sornette (2001) has shown that the correlation between different term of the string takes the form of

$$c(x,y) = \int_0^\infty dz g(x,z) g(y,z) \tag{2.15}$$

where g is some green function satisfying $\int_0^\infty dz[g(x,z)]^2 = 1,\ \forall x \geq 0$. They have shown that such a function $c(x,y)$ satisfies all the conditions for a correlation function and also that any correlation can be written as $c(x,y)$ with some suitable $g(x,z)$ function.

Some Parametric Examples

Case 1: If we assume the form $g(x,z) = \sqrt{J(x)} I_z[0, J(x)]$, where $J(x) > 0, \forall x$. Then we can have

$$c(x,y) = \sqrt{\frac{J(x) \wedge J(y)}{J(x) \vee J(y)}}.$$

For example,

- $J(x) = x$, so that $c(x,y) = \sqrt{\frac{x \wedge y}{x \vee y}} = e^{\frac{1}{2}|log(x) - log(y)|}$
- $J(x) = e^{2\kappa x^a}$, so that $c(x,y) = e^{\kappa |x^a - y^a|}$

Case 2: If we take $g(x,z)^2$ to be a probability density function of z, for example:

- $g(x,z)^2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-a(x))^2}{2}}$, then $c(x,y) = e^{-\frac{1}{8}(a(x)-a(y))^2}$

These are all legitimate correlation function of the string shock.

Some Comments

1. Finite factor HJM is different from general random field model, in terms of hedging strategy;

2. Compare multi-factor short rate model with random field model, if we impose the constrain of ***Condition 2*** on the volatility operator then we can only produce one-factor short rate model, but for general volatility, we can have infinite dimensional short rate as well.

2.4 The Volatility Dynamics

A widely used general functional form for volatility is like:

$$\sigma(t,T) = g(t,T)\phi(f(t,T))(V(t))^{\gamma}.$$

where $g(t,T)$ and ϕ are some deterministic functions; γ is usually either 0 or $\frac{1}{2}$; and $V(t,T)$ is a state variable follows some diffusion process. This volatility form will produce Markovian forward rate curves, i.e. future movements for forward rate curve only depend on the current curve and maybe some additional random components. Some special cases will be discussed in the following subsections.

2.4.1 Deterministic

The simplest characterization of the volatility process will be assuming it to be a deterministic function of current time t and time to maturity T, which equals the case of $\phi(x) = 1$ and $\gamma = 0$. This will result in a Gaussian model in t. Empirical evidence has supported a hump shaped volatility structure along the time to maturity direction, and this hump shape has been consistently observed over time. To accommodate this fact, Rebonnato (1999) has suggested the use of a four parameters parametric model for the volatility $\sigma(t,T) = g(T-t)$, which results in a special case of the one-factor HJM model.

The advantage of deterministic volatility is that we can always have analytical solution for European call option (Cap). In fact, Kennedy (1994) has given the formula of the Cap price under the general Gaussian random field case. What is more, we can also match the initial term structure of volatility perfectly, and it can also produce hump shaped volatility structure as time passes by. However, there are a couple of disadvantages too. For example, as time goes by even though it maintains a hump shape, it won't match the new volatility structure. That means it will have time dependent parameters which need to be constantly re-calibrated to match market volatility data. Another disadvantage is that the implied volatility will be flat as function of different strike prices. This is not consistent with the empirical evidence that smile or skewed shaped implied volatility curves are observed almost all the time. One of the solutions to this latter problem will be introducing a stochastic volatility as we will introduce in a later section.

Markovian Short Rate

Often times, we would like to price complex interest rate derivatives. One way of pricing them is through the construction of short rate tree. However, for a general volatility model, the corresponding tree structure is not necessarily recombining, which will render very complex tree. There is a subset of deterministic volatility models where the short rate process is Markovian, and thus a recombining tree can be constructed.

From

$$r(t) = f(t,t) = \int_0^t \sigma(u,t) \int_u^t \sigma(u,s) ds du + \int_0^t \sigma(s,t) dW(s),$$

it can be seen that with the specification that $\sigma(t,T) = \xi(t)\psi(T)$, the short rate process will be

$$r(t) = f(0,t) + \psi(t)\int_0^t \xi^2(u)\int_u^t \psi(s)dsdu + \psi(t)\int_0^t \xi(s)dW(s).$$

Note that in the one-factor case, it will degenerate to the general short rate model proposed by Hull and White(1990). Since if we define the deterministic function A(t) by:

$$A(t) := f(0,t) + \psi(t) \in_0^t \xi^2(u)\int_u^t \psi(s)dsdu.$$

then we will have the familiar form

$$dr(t) = [A'(t) + \psi'(t)\frac{r(t) - A(t)}{\psi(t)}]dt + \psi(t)\xi(t)dW(t) = [a(t) + b(t)r(t)]dt + c(t)dW(t).$$

Ritchken and Sankarasubramanian(1995) Framework

This section will generalized the above Markovian Short Rate process to the case that short rate may not necessarily Markovian, but it maybe a component of a higher-dimensional Markovian process, and thus a recombining lattice tree can be built in terms of this higher-dimensional process. This is proven in Ritchken and Sankarasubramanian (1995) and is rephrased here:

For a one-factor HJM model, if the volatility function $\sigma(t,T)$ is differentiable w.r.t. T, a necessary and sufficient condition for the price of any interest rate derivatives to be completely determined by a two state Markovian process $(r(.), \phi(.))^T$ is that the following condition holds:

$$\sigma(t,T) = \sigma_{RS}(t,T) := \eta(t)e^{-\int_t^T \kappa(x)dx},$$

where η is an adapted process and κ is an integrable deterministic function. In such a case, $\phi(t)$ will be

$$\phi(t) = \int_0^t \sigma_{RS}(s,t)ds.$$

and zero-coupon bond price will be given by

$$P(t,T) = \frac{P(0,T)}{P(0,t)} exp\{-\frac{1}{2}\Lambda^2(t,T)\phi(t) + \Lambda(t,T)[f(0,t) - r(t)]\}$$

where $\Lambda(t,T) = \int_t^T e^{-\int_t^u \kappa(x)dx} du$.

Differentiate this vector, it can be seen that under this RS class of volatility, short rate r evolves according to

$$d\begin{pmatrix} r(t) \\ \phi(t) \end{pmatrix} = \begin{pmatrix} \mu(r,t) \\ \eta^2(t) - 2\kappa(t)\phi(t) \end{pmatrix} dt + + \begin{pmatrix} \eta(t) \\ 0 \end{pmatrix} dW(t)$$

with

$$\mu(r,t) = \kappa(t)[f(0,t) - r(t)] + \phi(t) + \frac{\partial}{\partial t} f(0,t).$$

From these, we can see that η is just the instantaneous short rate volatility process.

2.4.2 Affine

Following affine factor model for Bond prices (Heston 1993), Collin-Dufresne and Goldstein (2003) has extended the affine term structure model to the random field model (i.e. infinite factor model). By modeling the log-Bond prices to follow a random field model with CIR type Factor volatility process, they produce analytical solution to a wide range of derivatives, just like the finite factor affine model. These are achieved by restricting the characteristic function of bond price process to be analytical, through an affine factor CIR type dynamics for the innovation of volatility.

Suppose we assume that the risk-neutral zero-coupon bond price to follow the process that:

$$\frac{dP(s,T)}{P(s,T)} = r(s)ds - \sigma(s,T)\sqrt{\Sigma(s)}d\mathbf{W}(s,T) \tag{2.16}$$

where $\sigma(s,T)$ is an arbitrary deterministic function, $\mathbf{W}(s) = \{\mathbf{W}(s,T), s \leq T \leq T_{Max}\}$ is an infinite dimensional Wiener process with correlation kernel $c(,)$ and the volatility state variable $\Sigma(s)$ follows

$$d\Sigma(s) = \kappa(\theta - \Sigma)ds + \vartheta\sqrt{\Sigma}dB(s) \tag{2.17}$$

where $B(s)$ is a finite-factor Brownian motion independent of the Wiener process $\mathbf{W}(s)$.

Collin-Dufresne and Goldstein (2003) has shown that under some regularity conditions on the parameters of the model, the characteristic function of future log-bond prices will be log-affine in terms of log-bond prices and volatility state variable of current time. Let

$$\psi_t(\alpha) := E_t^Q[e^{\alpha_0\Sigma(T_0)+\sum_{j=1}^n \alpha_j log P^{T_j}(T_0)}] \tag{2.18}$$

$$\psi_t^0(\alpha) := E_t^Q[e^{-\int_t^{T_0} r_s ds}e^{\alpha_0\Sigma(T_0)+\sum_{j=1}^n \alpha_j log P^{T_j}(T_0)}] \tag{2.19}$$

Then it can be shown that:

$$\psi_t(\alpha) = exp\left(M(t) + N(t)\Sigma(t) + \sum_{j=1}^n \alpha_j log\frac{P^{T_j}(t)}{P^{T_0}(t)}\right)$$

and

$$\psi_t^0(\alpha) = P^{T_0}(t)exp\left(M^0(t) + N^0(t)\Sigma(t) + \sum_{j=1}^n \alpha_j log\frac{P^{T_j}(t)}{P^{T_0}(t)}\right)$$

Where the deterministic functions M, N, M_0, N_0 will be determined by a system of ODE's.

Note that, modeling Bond prices to follow Random Field process does not guarantee that forward rate will follow a random field. However, by specifying the dynamics of Bond prices, we would have automatically determined the dynamics of the forward rate process.

2.4.3 Log-Gaussian Form

To ensure volatility to be always positive, another type of popular yet natural choice of volatility is in the log form. In other word, instead of modeling volatility directly, we take the log first and then model the log-volatility (as Gaussian) without the worry of it being negative. A number of researchers have followed this route in their modeling of both equity and fixed income products, e.g. Cheng et al (2008). Here we will follow this log-formed formulation as well. Beside the requirement of forward rate to follow random field process, we will model the log-volatility to follow Gaussian random field as well:

$$
\begin{aligned}
&log\left(\sigma^2(t,T)\right) = h(t,T) \\
&h(t,T) = [\alpha(t,T) - \beta(t,T)\ h(t,T)]\ dt + \sigma_h(t,T)\ d\mathbf{W}(t,T)
\end{aligned}
$$

The intuition is similar to the requirement on forward rate, since the forward rate and its volatility are both functions of their time to maturity. They should be looked at in a general Hilbert space, and thus shall be modeled as random field. In chapter 3 and 4, we will introduce some specific random field model on both forward rate and its volatility, and then propose an approximation method of pricing simple interest rate option under these specifications.

2.5 Market Completeness

2.5.1 Theory

A financial market place is said to be complete when a market exists with an equilibrium price for every asset in every possible state of the world.[2] The notion of market completeness has been closely related with hedging in the sense that, in a complete market any contingent claim can be replicated by a portfolio of assets that already exist in the market. When it comes to modeling, financial asset prices are usually modeled as diffusion processes driven by Wiener processes and there is an informal principle (Bensoussan 1984) that: to hedge against N sources of randomness one needs N non-redundant securities besides the numeraire. According to this principle, in a N-factor model the market can be completed by N non-redundant assets plus a bank account, and any contingent claims (e.g. options) can be replicated and thus priced by the N assets. In particular, for the bond market where there is a continuum of securities (for different maturity dates), one can construct hedging strategies involving a continuum of assets for any contingent claims. Nevertheless, Bjork et al (1997) has shown that in general if we assume infinite factor random sources, e.g. a function valued Wiener process or random field, one can hedge in the most favorable situation only a dense subset in the space of contingent claims, and thus proposed the concept of approximate completeness [3] as the fundamental concept. Various empirical findings in the literature provide examples of incomplete market. For instance, under the complete market assumption it is possible to price options (caps and swaptions) accurately from stocks (bonds); however, financial models calibrated based on stocks (bonds) alone cannot fit

[2] Definition quoted by OECD from IMF, 2003, External Debt Statistics: Guide for Compilers and Users – Appendix III, Glossary, IMF, Washington DC

[3] Every contingent claim can be approached in a certain sense by a sequence of hedgeable claims.

the option prices accurately, nor can they adequately explain the features exhibited by them. To match empirical results better, researchers have proposed certain refined models that result in incomplete markets with empirical evidences more consistent with the real market. One such a strategy is to model the random source as a Levy process with jumps added to the diffusive evolution. Another strategy is to introduce various stochastic volatility (SV) models [e.g. Hull and White (1987), Heston (1993), Scott (1997), etc.] that contain sources for randomness driving the volatility structure in addition to those driving asset prices directly. Both jump diffusion models and SV models create incomplete markets. SV models were introduced for bond markets, e.g. in Chiarella and Kwon (2000), Wiener processes drive both the forward rate process and the volatility process. Such an approach is analogous to those SV counterparts for the equity market.

Random field term structure models (string models) consist of infinite factors to which the traditional HJM framework is a special case with infinite states but finite factors. As was commented in Goldstein (2000), a random field model generally implies the existence of an infinite number of economic factors, and hence no riskless portfolio can be constructed if only bonds of different maturities are used. Thus, rather than identifying the risk-neutral measure, which is not unique in this setting, we assume its existence (as otherwise there would exist arbitrage opportunity in the market) and write down the dynamics of the forward rates as fundamental under this measure so that various fixed-income derivatives can be priced. Furthermore, the term structure model we adopt in this dissertation consists of two random fields (with or without correlations between them), one for the forward rate and the other for stochastic volatility. Therefore, it makes the market incomplete, the risk-neutral measure non-unique, and the non-identifiability issue inevitable. We have to follow the aforementioned approach in Goldstein (2000) to assume that the risk-neutral dynamics follow a certain martingale measure Q.

2.5.2 Empirical Studies

The main question in the studies of empirical option pricing is: Which risk-neutral measure gives rise to samples that look consistent with real data collected in the bond and related option markets? Such an issue may involve estimating model parameters, some of which correspond to the risk premia.

There is a large literature for calibration of SV models in equity markets using both asset returns and option data. E.g. Renault and Touzi (1996) showed the natural fit between implied volatility smile and SV model; Cheng et al (2008) used both stock and option prices to calibrate their model. Similar contributions were made to fit HJM finite factor models, e.g., Casassus, Collin-Dufresne and Goldstein (2005), Trolle and Schwartz (2009). In contrast, empirical studies in random field models are not well developed. There are plenty of challenges in this area. For one thing, bonds alone cannot complete the bond market. As remedies, Han (2007) included swaption data in his model; Jarrow, Li and Zhao (2007) calibrated their model with cap skew data; Trolle and Schwartz (2009) used both cap and swaption; Collin-Dufresne, Goldstein and Jones (2009) documented the so called unspanned stochastic volatility; Longstaff, Santa-Clara and Schwartz (2001) studied the relative valuation of cap and swaption and found that cap prices periodically deviate significantly from the no-arbitrage values implied by the swaption market and thus neither of them should be omitted from the model calibration. All of these implied that fixed-income derivatives such as caps and swaptions contain needed information to explain market behaviors that cannot be accomplished by using bonds alone.

Considering the non-identifiability issue regarding SV random field term structure models, and following the approximate completeness approach in Bjork et al (1997), it is necessary to resort to good approximation schemes when calibrating a SV random field term structure model. This amounts to dimension reduction in the sense of choosing finite factors and performing inference on the proxy. Such an approach is illustrated in Han (2007) and Trolle and Schwartz (2009), and will also be our direction of effort into the next phrase of model calibration. Han (2007) started under risk neutral measure with a random field model for Zero bond prices with random covariance, but then had to reduce dimension to finite factors by keeping the first several principal components for implementation. Furthermore, to bridge the risk neutral measure (for option pricing) with the physical measure (for estimation), he chose a special form of market price of risk process such that the structure of the bond price and volatility processes remain the same under both measures. This process of starting with models under risk neutral measure and then pick a convenient risk premium process represent a typical path of model calibration for a SV model, including Trolle and Schwartz (2009) which, as far as I am aware, is the most general factor SV HJM model that incorporates a wide range of option data.

Chapter 3

Option Pricing in Random Field Model with Random Field Volatility

Options on interest rate have been actively traded in the market, with interest rate cap beng one of the most popular. An interest rate cap is a derivative from which the buyer has the option to receive a series of payments at the end of each period, specified in the contract, during which the interest rate exceeds the agreed strike rate. E.g. an agreement to receive payments for each month the monthly LIBOR rate exceeds 2.5% for the next 12 months. A cap usually consists of a series of such options for successive time periods, but it suffices to consider just one period which is called a caplet, since the payoff or price for a cap will simply be the sum of these one period caplets. In this chapter, we will study the pricing of cap under random field forward rate model with random field volatility. For simplicity, what we call cap in the rest of the dissertation will only be for one period and thus essentially is a caplet.

To price a cap, we need first to look at the payoff at the maturity date of the cap, which will depend on the forward rate curve at that moment. Now let's look at it mathematically: $\forall 0 \leq s \leq t$, let $f(s,t)$ be the instantaneous forward rate at time s

with maturity of time t. For $0 \leq s \leq t < t + \Delta \leq T$, let

$$f^{\Delta}(s,t) = \Delta^{-1} \int_{t}^{t+\Delta} f(s,u)\, du \tag{3.1}$$

then $f^{\Delta}(s,t)$ will be the effective interest rate that can be locked in at time s and effective for the period $[t, t+\Delta]$. Consider an interest rate cap with strike rate K for the period $[t, t+\Delta]$, which can be regarded as a European option on $f^{\Delta}(s,t)$, the holder will exercise the option at time t if $f^{\Delta}(t,t) > K$, this will yield a payoff at $t + \Delta$ of

$$\left((e^{\Delta\ f^{\Delta}(t,t)} - 1) - (e^{\Delta K} - 1) \right)^{+} = \left(e^{\Delta\ f^{\Delta}(t,t)} - e^{\Delta K} \right)^{+}.$$

From the fundamental theorem of asset pricing, the cap price at time 0 will be the expected discounted payoff of the option under the risk-neutral measure:

$$E^{Q} \left\{ \exp\left(- \int_{0}^{t+\Delta} r(u)\, du \right) \left(e^{\Delta\ f^{\Delta}(t,t)} - e^{\Delta K} \right)^{+} \right\}. \tag{3.2}$$

In this chapter, we will discuss the pricing of cap under a random field forward rate model with random field volatility. We start with a random field forward rate model with deterministic volatility in section 1, under which Kennedy (1994) has derived a closed form formula for cap price. Utilizing his result and the law of iterated expectation, in section 2 we will show that cap price under a random field volatility will simply be an expected price conditioning on the past and the future movements of volatility. Then we will propose an approximation scheme in section 3 to facilitate the calculation of the expectation. After that in section 4, we will use simulation to study the accuracy and efficiency of this approximation and further talk about the effects of some model parameters. Note that, in this chapter we will assume independent noises between forward rates and their volatilities. The correlated case, which might be more realistic in real market, is a much more complex problem and will be discussed in the next chapter.

3.1 Option Pricing under Gaussian Random Field Forward Rate Model

Under Gaussian Random Field Model, i.e. deterministic volatility structure, Kennedy (1994) has derived a closed form formula for Caplet, which is essential an European style put option on zero coupon bond. Here we briefly rewrite his result:

Under the settings of last chapter and also the ***Condition 1-2***, we consider a random field model on forward rate under risk-neutral measure as

$$f(t,T) = \int_0^t \mu(s,T)ds + \int_0^t \sigma(s,T)d\mathbf{W}(s,T)$$

where

- $\mathbf{W}(s)(\cdot) = \{\mathbf{W}(s, s+u), 0 \le u \le T_{max}\}$ is a Wiener process taking values in the Hilbert space $H = C[0, T_{max}] = \{$All continuous functions on $[0, T_{max}]$ $\}$ with deterministic kernel function $c(\cdot,\cdot)$ for its covariance operator.

- $f(t,T) = \mathbf{f}(t)(T-t)$ and $\mathbf{W}(t,T) = \mathbf{W}(t)(T-t)$, where T is the maturity time, $0 \le t \le T \le t + T_{max}$ and $T - t$ is the time to maturity.

- $d\mathbf{W}(s,u)\ d\mathbf{W}(s,v) = c(u,v)ds$ and $d\mathbf{W}(s,u)\ d\mathbf{W}(t,v) = 0,\ \forall s \ne t.$

- $\sigma(s,T)$ is deterministic, and $\mu(s,T)$ satisfies the no-arbitrage condition

$$\mu(s,T) = \sigma(s,T)\int_s^T \sigma(s,u)c(T,u)du$$

.

Corollary 5. *The time* 0 *price of a cap with strike price* K *between the period* $[t, t+\Delta]$ *from expectation (3.2) is given by*

$$\begin{aligned} & C(0, \sigma_t^2) \\ = \; & \exp\left[-\int_0^{t+\Delta} f(0,u)\ du\right] \left\{\exp\left[\int_t^{t+\Delta} f(0,u)\ du\right] \ \Phi(b(\sigma_t) + \sigma_t/2) \right. \\ & \left. - \ e^{\Delta K} \ \Phi(b(\sigma_t) - \sigma_t/2)\right\} \end{aligned} \tag{3.3}$$

where Φ *is the cdf of* $N(0,1)$ *distribution,* $b(\sigma_t) = \left[\int_t^{t+\Delta} f(0,u)\ du - \Delta K\right] \ \sigma_t^{-1}$*, and the variance*

$$\begin{aligned} \sigma_t^2 &= Var\left(\int_t^{t+\Delta} f(t,u)\ du\right) \\ &= \int_t^{t+\Delta}\int_t^{t+\Delta}\int_0^t \sigma(u,v_1)\sigma(u,v_2)c(v_1,v_2)\ du\ dv_1\ dv_2. \end{aligned} \tag{3.4}$$

Proof: Let $N_1 = \Delta f^{\Delta}(t,t) = \int_t^{t+\Delta} f(t,u)du$ and $N_2 = \int_0^{t+\Delta} R_u du = \int_0^{t+\Delta} f(u,u)du$ then

$$N1 = \int_t^{t+\Delta} f(0,u)du + \int_t^{t+\Delta}\int_0^t \mu(u,v)dudv + \int_t^{t+\Delta}\int_0^t \sigma(u,v)d_u\mathbf{W}(u,v)dv,$$
$$N2 = \int_0^{t+\Delta} f(0,u)du + \int_0^{t+\Delta}\int_0^v \mu(u,v)dudv + \int_0^{t+\Delta}\int_0^v \sigma(u,v)d_u\mathbf{W}(u,v)dv$$

and $(N_1, N_2)^T$ is a bivariate normal vector. The price of the cap from the expectation (3.2) can be shown as:

$$\begin{aligned} E\left[e^{-N_2}(e^{N_1} - e^{\Delta K})^+\right] \;=\; & e^{E(N_1-N_2)+\frac{Var(N_1-N_2)}{2}} \ \Phi\left(\frac{E(N_1) + Var(N_1) - Cov(N_1,N_2) - \Delta K}{\sqrt{Var(N_1)}}\right) \\ & -e^{\Delta K - E(N_2) + \frac{Var(N_2)}{2}} \ \Phi\left(\frac{E(N_1) - Cov(N_1,N_2) - \Delta K}{\sqrt{Var(N_1)}}\right) \end{aligned} \tag{3.5}$$

From the no arbitrage drift condition and the assumption that $d\mathbf{W}(u_1, v_1)\, d\mathbf{W}(u_2, v_2) = c(v_1, v_2)I_{\{u_1=u_2\}}du_1$, we can obtain the mean, variance and covariance between N_1 and N_2:

$$
\begin{aligned}
EN_1 &= \int_t^{t+\Delta} f(0,u)du + \int_t^{t+\Delta}\int_0^t \sigma(u,v_1)\int_u^{v_1}\sigma(u,v_2)c(v_1,v_2)dv_2dudv_1 \\
EN_2 &= \int_0^{t+\Delta} f(0,u)du + \int_0^{t+\Delta}\int_0^{v_1} \sigma(u,v_1)\int_u^{v_1}\sigma(u,v_2)c(v_1,v_2)dv_2dudv_1
\end{aligned}
$$

$$
\begin{aligned}
Var(N_1) &= \int_t^{t+\Delta}\int_t^{t+\Delta}\int_0^t \sigma(u,v_1)\sigma(u,v_2)c(v_1,v_2)dudv_1dv_2 \\
Var(N_2) &= \int_0^{t+\Delta}\int_0^{t+\Delta}\int_0^{v_1\wedge v_2} \sigma(u,v_1)\sigma(u,v_2)c(v_1,v_2)dudv_1dv_2 \\
Cov(N_1,N_2) &= \int_0^{t+\Delta}\int_t^{t+\Delta}\int_0^{v_1\wedge t} \sigma(u,v_1)\sigma(u,v_2)c(v_1,v_2)dudv_1dv_2
\end{aligned}
$$

Substitute the above results into (3.5) will give equation (3.3). *QED*.

3.2 Random Field Forward Rate with Random Field Volatility

Instead of being deterministic if we assume volatility to follow a random field process, the forward rate process under the risk-neutral measure Q satisfies

$$
\begin{aligned}
df(t,T) &= \mu(t,T)\, dt + \sigma(t,T)\, d\mathbf{W}_1(t,T), \qquad (3.6) \\
d\ log\sigma^2(t,T) &= [\alpha - \beta\ log\sigma^2(t,T)]\, dt + \sigma_h\, d\mathbf{W}_2(t,T), \qquad (3.7)
\end{aligned}
$$

where α, β and $\sigma_h \in H$ are continuous deterministic functions of time and time to maturity and $\{\mathbf{W}_i(t, \cdot)\}$, $i = 1, 2$ are two independent Wiener processes with covariance kernels $c_i(\cdot, \cdot)$, respectively, that satisfy the following correlation structures:

(i) $d\mathbf{W}_j(t, T_1)\ d\mathbf{W}_j(t, T_2) = c_j(T_1, T_2)\ dt, \quad j = 1, 2,\ \forall\ t \leq \min\{T_1, T_2\}$; in particular, $c_j(T, T) = 1,\ j = 1, 2$.

(ii) $d\mathbf{W}_j(s, T_1)\ d\mathbf{W}_j(t, T_2) = 0, \quad j = 1, 2,\ \forall s \neq t$, with $\max\{s, t\} \leq \min\{T_1, T_2\}$.

(iii) $\{\mathbf{W}_1(t, T)\}$ and $\{\mathbf{W}_2(t, T)\}$ are independent.

Note that (3.6) — (3.7) form a SV model with white noise shocks. But pricing bond derivatives may become much more difficult with this model for a number of reasons. Consider an European call option, written on the zero-coupon bond, with maturity $\tau \in [t, T]$ and strike price K. Its time t price can be formally expressed as

$$C(t, \tau, T) = E_t^Q \left\{ \exp\left[-\int_t^\tau r(u) du \right]\ [P(\tau, T) - K]^+ \right\}. \tag{3.8}$$

First, the forward rate process $\{f(u, T)\}$ will determine the short rate $\{r(u)\}$ and the bond price $\{P(u, T)\}$ in (3.8). Second, no generalized Black-Scholes (GBS) formulas are available for computing $C(t, \tau, T)$. In theory, the conditional expectation in (3.8) is still a high-dimensional integral and could be calculated via brute force Monte Carlo simulation. But this is totally impractical when such numerical integration needs to be performed repeatedly in an iterative calibration algorithm. We propose to approximate $C(t, \tau, T)$ based on an expectation over a low-dimensional (joint) density. With an adequate proxy density, we can work out certain cases with specific correlation functions in the model (3.6) — (3.7). The reason for us to study the model (3.6) — (3.7) is that it has similar features to some previously studied SV models, and it does not entertain a closed-form solution for bond pricing by using the Fourier inversion method introduced in Heston (1993). Therefore, bond pricing with this model certainly requires some novel probability approximation schemes.

The forward rates $\{f(t,T)\}$ in Kennedy (1994) form a Gaussian field. Following (3.6), we have

$$f(t,T) \;=\; f(0,T) + \int_0^t \mu(u,T)\;du + \int_0^t \sigma(u,T)\;d\mathbf{W}_1(u,T) \tag{3.9}$$

which is not Gaussian. Instead, $\{f(t,T)\}$ conditioning on $\{\sigma(t,T)\}$ is a Gaussian field with mean

$$m(t,T) = f(0,T) + \int_0^t \mu(u,T)\;du$$

variance

$$V(t,T) = \int_0^t \sigma^2(u,T)\;du,$$

and covariance

$$\begin{aligned}
& c(t_1 \wedge t_2, T_1, T_2) \triangleq Cov(f(t_1,T_1), f(t_2,T_2)) \\
= \;& E_h\left\{\int_0^{t_1\wedge t_2} \sigma(u,T_1)\;d\mathbf{W}_1(u,T_1)\cdot \int_0^{t_1\wedge t_2} \sigma(u,T_2)\;d\mathbf{W}_1(u,T_2)\right\} \\
= \;& \int_0^{t_1\wedge t_2} \sigma(u,T_1)\sigma(u,T_2)\;c_1(T_1,T_2)\;du,
\end{aligned}$$

where $E_h(\cdot)$ denotes the conditional expectation [of the white noise processes $\{\mathbf{W}_i(t,T)\}$, $i=1,2$] given the stochastic volatility $\{\sigma(t,T)\}$.

Condition for no arbitrage

The No-arbitrage condition in this case from (2.3) will be:

$$\mu(s,T) = \sigma(s,T)\int_s^T \sigma(s,u)c_1(T,u)du \tag{3.10}$$

3.2.1 Cap Price Formula

Here let's see how to price option under this more general framework. Conditioning on the history of volatility structure, our current model will be Gaussian. Making use of the result of Corollary 5, the time 0 price of the above cap is given by

$$
\begin{aligned}
& C(0,\sigma_t^2) \\
= \ & \exp\left[-\int_0^{t+\Delta} f(0,u)\ du\right] \left\{\exp\left[\int_t^{t+\Delta} f(0,u)\ du\right]\ \Phi(b(\sigma_t)+\sigma_t/2)\right. \\
& \quad \left. -\ e^{\Delta K}\ \Phi(b(\sigma_t)-\sigma_t/2)\right\}, \qquad (3.11)
\end{aligned}
$$

where Φ is the cdf of $N(0,1)$ distribution, $b(\sigma_t) = \left[\int_t^{t+\Delta} f(0,u)\ du - \Delta K\right]\ \sigma_t^{-1}$, and the conditional variance

$$
\begin{aligned}
\sigma_t^2 &= Var\left(\int_t^{t+\Delta} f(t,u)\ du \mid h(s,t):\ 0 \le s \le t \le T\right) \\
&= 2\int_t^{t+\Delta}\int_t^{v_2}\int_0^t \sigma(u,v_1)\sigma(u,v_2)\ c_1(v_1,v_2)\ du\ dv_1\ du_2. \qquad (3.12)
\end{aligned}
$$

3.2.2 A Probability Approximation Scheme for the Distribution of Summary Statistic σ_t^2

(3.6) and (3.7) differ from Kennedy's model due to the presence of the stochastic volatility factor in (3.7). To extend the pricing formula (3.3), a natural attempt is to define the cap price by

$$C(0) = E[C(0,\sigma_t^2)], \qquad (3.13)$$

where the expectation is taken over σ_t^2 under the risk neutral measure and can be computed via Monte Carlo numerical integration. Note that σ_t^2 is a function of latent volatility field $\{\sigma(t,T)\}$. The standard brute force simulation is to generate a large number of future volatility sample paths, calculate σ_t^2 and $C(0,\sigma_t^2)$ over each path, then take a sample average as a proxy of $C(0)$. Such a method would work if pricing is the ultimate goal. However, in model calibration and related problems, derivative prices have to be calculated repeatedly at every site (t, T) and in every iteration of an iterative algorithm (e.g. MCMC). Alternative methods to the brute force simulation are clearly needed to alleviate the computational intensity. See Cheng et al. (2008) for a recently proposed Gaussian approximation scheme. In this paper, we follow the same idea but propose to adopt a log-normal approximate distribution for σ_t^2. Here is an outline of the proposed procedure:

Step 1 Compute the first and second moments of σ_t^2. Certain low-dimensional numerical integration is required in this step.

Step 2 Convert the two moments to the two parameters (η_1, η_2) in log-normal distribution.

Step 3 Compute the expectation $E[C(0, \sigma_t^2)]$ with respect to the approximate log-normal density obtained in Step 2, hence get an approximate price $C(0)$.

The appeal of this approach is to reduce the dimensionality in Monte Carlo from at least several hundreds (the total number of components of the volatility field $\{h(t, T)\}$) to one (a single log-normal random variable) and still maintain the accuracy to a satisfactory degree. In what follows, we provide details in those three steps.

3.2.3 Moments of σ_t^2

The log-volatility as the solution to (3.7) is an Ornstein-Uhlenbeck (O-U) process

$$log\sigma^2(t, T) = e^{-\int_0^t \beta(s,T)ds} log\sigma^2(0, T) + \int_0^t e^{-\int_u^t \beta(s,T)ds} \alpha(u, T)du + \int_0^t e^{-\int_u^t \beta(s,T)ds} \sigma_h(u, T)dW_2(u, T). \tag{3.14}$$

For fixed v_1 and v_2, $\forall\ u \leq v_1 \leq v_2$ the random exponent in (3.12),

$$\begin{aligned}
& g^{v_1,v_2}(u) \triangleq \frac{1}{2}\ [log\sigma^2(u,v_1) + log\sigma^2(u,v_2)] \\
= & \ \frac{1}{2}\left[e^{-\int_0^u \beta(s,v_1)ds} log\sigma^2(0,v_1) + e^{-\int_0^u \beta(s,v_2)ds} log\sigma^2(0,v_2)\right] \\
& +\frac{1}{2}\left[\int_0^u e^{-\int_\tau^u \beta(s,v_1)ds}\alpha(\tau,v_1)d\tau + \int_0^u e^{-\int_\tau^u \beta(s,v_2)ds}\alpha(\tau,v_2)d\tau\right] \\
& +\frac{1}{2}\left[\int_0^u e^{-\int_\tau^u \beta(s,v_1)ds}\sigma_h(\tau,v_1)dW_2(\tau,v_1) + \int_0^u e^{-\int_\tau^u \beta(s,v_2)ds}\sigma_h(\tau,v_2)dW_2(\tau,v_2)\right]
\end{aligned}$$

is a Gaussian process in index u with mean

$$\begin{aligned}
E\left[g^{v_1,v_2}(u)\right] &= \frac{1}{2}\left[e^{-\int_0^u \beta(s,v_1)ds} log\sigma^2(0,v_1) + e^{-\int_0^u \beta(s,v_2)ds} log\sigma^2(0,v_2)\right] \\
&\quad +\frac{1}{2}\left[\int_0^u e^{-\int_\tau^u \beta(s,v_1)ds}\alpha(\tau,v_1)d\tau + \int_0^u e^{-\int_\tau^u \beta(s,v_2)ds}\alpha(\tau,v_2)d\tau\right],
\end{aligned}$$

variance

$$\begin{aligned}
Var\left[g^{v_1,v_2}(u)\right] = \frac{1}{4}\Bigg[&\int_0^u e^{-2\int_\tau^u \beta(s,v_1)ds}\sigma_h^2(\tau,v_1)d\tau + \int_0^u e^{-2\int_\tau^u \beta(s,v_2)ds}\sigma_h^2(\tau,v_2)d\tau \\
&+2\int_0^u e^{-\int_\tau^u [\beta(s,v_1)+\beta(s,v_2)]ds}\sigma_h(\tau,v_1)\sigma_h(\tau,v_2)c_2(v_1,v_2)d\tau\Bigg],
\end{aligned}$$

and covariance (for $u_1 \leq v_{11} \leq v_{21}$ and $u_2 \leq v_{12} \leq v_{22}$)

$$\begin{aligned}
& Cov\left(g^{v_{11},v_{21}}(u_1),\ g^{v_{12},v_{22}}(u_2)\right) \\
= & \ \frac{1}{4}\int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{11})+\beta(s,v_{12})]ds}\sigma_h(\tau,v_{11})\sigma_h(\tau,v_{12})c_2(v_{11},v_{12})d\tau \\
& +\frac{1}{4}\int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{11})+\beta(s,v_{22})]ds}\sigma_h(\tau,v_{11})\sigma_h(\tau,v_{22})c_2(v_{11},v_{22})d\tau \\
& +\frac{1}{4}\int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{12})+\beta(s,v_{21})]ds}\sigma_h(\tau,v_{12})\sigma_h(\tau,v_{21})c_2(v_{12},v_{21})d\tau \\
& +\frac{1}{4}\int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{12})+\beta(s,v_{22})]ds}\sigma_h(\tau,v_{12})\sigma_h(\tau,v_{22})c_2(v_{12},v_{22})d\tau
\end{aligned}$$

Next we derive the first and second moments of σ_t^2, based on which numerical computation of $C(0)$ will be carried out.

Proposition 1. *Under the setting of (3.6) – (3.7) , we have*

$$
\begin{aligned}
E\sigma_t^2 &= 2 \int_t^{t+\Delta} \int_t^{v_2} c_1(v_1, v_2) \int_0^t \exp[H_1(u, v_1, v_2)]\ du\ dv_1\ dv_2 && (3.15) \\
E(\sigma_t^2)^2 &= 4 \int_t^{t+\Delta} \int_t^{t+\Delta} \int_t^{v_{22}} \int_t^{v_{21}} c_1(v_{11}, v_{21})\ c_1(v_{12}, v_{22}) && (3.16) \\
& \qquad \int_0^t \int_0^t \exp[H_2(u_1, u_2, v_{11}, v_{12}, v_{21}, v_{22})]\ du_1\ du_2\ dv_{11}\ dv_{12}\ dv_{21}\ dv_{22},
\end{aligned}
$$

where for $u \leq v_1 \leq v_2$, $u_1 \leq v_{11} \leq v_{21}$ and $u_2 \leq v_{12} \leq v_{22}$ we have

$$
\begin{aligned}
& H_1(u, v_1, v_2) = E\left[g^{v_1, v_2}(u)\right] + \frac{1}{2}\ Var\left[g^{v_1, v_2}(u)\right] && (3.17) \\
= \ & \frac{1}{2}\left[e^{-\int_0^u \beta(s, v_1) ds} log\sigma^2(0, v_1) + e^{-\int_0^u \beta(s, v_2) ds} log\sigma^2(0, v_2)\right] \\
& + \frac{1}{2}\left[\int_0^u e^{-\int_\tau^u \beta(s, v_1) ds} \alpha(\tau, v_1) d\tau + \int_0^u e^{-\int_\tau^u \beta(s, v_2) ds} \alpha(\tau, v_2) d\tau\right] \\
& + \frac{1}{8}\left[\int_0^u e^{-2\int_\tau^u \beta(s, v_1) ds} \sigma_h^2(\tau, v_1) d\tau + \int_0^u e^{-2\int_\tau^u \beta(s, v_2) ds} \sigma_h^2(\tau, v_2) d\tau\right] \\
& + \frac{1}{4} \int_0^u e^{-\int_\tau^u [\beta(s, v_1) + \beta(s, v_2)] ds} \sigma_h(\tau, v_1) \sigma_h(\tau, v_2) c_2(v_1, v_2) d\tau
\end{aligned}
$$

and

$$
\begin{aligned}
& H_2(u_1, u_2, v_{11}, v_{12}, v_{21}, v_{22}) \qquad (3.18)\\
= \;& E\left[g^{v_{11},v_{21}}(u_1)\right] + E\left[g^{v_{12},v_{22}}(u_2)\right] + \frac{1}{2}\; Var\left[g^{v_{11},v_{21}}(u_1)\right] + \frac{1}{2}\; Var\left[g^{v_{12},v_{22}}(u_2)\right]\\
& \quad + Cov\left(g^{v_{11},v_{21}}(u_1),\; g^{v_{12},v_{22}}(u_2)\right)\\
= \;& H_1(u_1, v_{11}, v_{21}) + H_1(u_2, v_{12}, v_{22})\\
& + \frac{1}{4} \int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{11})+\beta(s,v_{12})]ds} \sigma_h(\tau, v_{11})\sigma_h(\tau, v_{12}) c_2(v_{11}, v_{12}) d\tau\\
& + \frac{1}{4} \int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{11})+\beta(s,v_{22})]ds} \sigma_h(\tau, v_{11})\sigma_h(\tau, v_{22}) c_2(v_{11}, v_{22}) d\tau\\
& + \frac{1}{4} \int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{12})+\beta(s,v_{21})]ds} \sigma_h(\tau, v_{12})\sigma_h(\tau, v_{21}) c_2(v_{12}, v_{21}) d\tau\\
& + \frac{1}{4} \int_0^{u_1\wedge u_2} e^{-\int_\tau^{u_1\wedge u_2}[\beta(s,v_{12})+\beta(s,v_{22})]ds} \sigma_h(\tau, v_{12})\sigma_h(\tau, v_{22}) c_2(v_{12}, v_{22}) d\tau
\end{aligned}
$$

Proof: The fact that $g^{v_1,v_2}(u)$ is Gaussian implies

$$
\begin{aligned}
E\sigma_t^2 &= 2 \int_t^{t+\Delta} \int_t^{v_2} c_1(v_1, v_2) \int_0^t E\left\{\exp[g^{v_1,v_2}(u)]\right\}\; du dv_1 dv_2\\
&= 2 \int_t^{t+\Delta} \int_t^{v_2} c_1(v_1, v_2) \int_0^t \exp\{Eg^{v_1,v_2}(u) + Var[g^{v_1,v_2}(u)]/2\}\; du dv_1 dv_2.
\end{aligned}
$$

Hence (3.15) and (3.17) follow from (3.15) and (3.15). By the same token, (3.16) and (3.18) follow from (3.15), (3.15) and (3.15). QED.

3.3 Numerical Study

In this section, we perform simulation study to check the speed and accuracy of the proposed approximation method. Also, we will examine the effect of model parameters. For simplicity, we will consider α, β and σ_h to be constants instead of functions in this section.

3.3.1 Simulation of the Distribution of σ_t^2 by Monte Carlo Method

since

$$\sigma_t^2 = Var(\int_t^{t+\delta} f^u(t)du) = 2\int_t^{t+\delta} du \int_t^u \int_0^t \sigma(s,u)\sigma(s,\nu)c_1(u,\nu)dsd\nu$$

where

$$log\sigma^2(t,T) = e^{-\beta t}log\sigma^2(0,T) + \frac{\alpha}{\beta}(1-e^{-\beta t}) + \sigma e^{-\beta t}\int_0^t e^{\beta s}d\mathbf{W}_2(s,T)$$

Then

$$\begin{aligned} g^{u,\nu}(s) &= (log\sigma^2(s,u) + log\sigma^2(s,\nu))/2 \\ &= \frac{\alpha}{\beta} + e^{-\beta s}(\frac{log\sigma^2(0,u)+log\sigma^2(0,\nu)}{2} - \frac{\alpha}{\beta}) + \sigma e^{-\beta s}\frac{\int_0^s e^{\beta y}d\mathbf{W}_2(y,u) + \int_0^s e^{\beta y}d\mathbf{W}_2(y,\nu)}{2} \\ &= \frac{\alpha}{\beta} + e^{-\beta s}(\frac{log\sigma^2(0,u)+log\sigma^2(0,\nu)}{2} - \frac{\alpha}{\beta}) + \sigma e^{-\beta s}X^{u,\nu}(s) \\ &:= a(s,u,v) + b(s)X^{u,\nu}(s) \end{aligned} \tag{3.19}$$

Where

$$a(s,u,v) = \frac{\alpha}{\beta} + e^{-\beta s}(\frac{log\sigma^2(0,u)+log\sigma^2(0,\nu)}{2} - \frac{\alpha}{\beta})$$

$$b(s) = \sigma e^{-\beta s}$$

and

$$X^{u,\nu}(s) = \frac{\int_0^s e^{\beta y}d\mathbf{W}_2(y,u) + \int_0^s e^{\beta y}d\mathbf{W}_2(y,\nu)}{2}$$

satisfies $X^{u,\nu}(s+ds) - X^{u,\nu}(s) \perp X^{u,\nu}(s)$ and $X^{u,\nu}(s+ds) - X^{u,\nu}(s) \sim N(0, \Sigma_s)$ with

$$\{\Sigma_s\}_{i,j} = [c_2(u_i, v_i) + c_2(u_i, v_j) + c_2(v_i, u_j) + c_2(v_i, v_j)] \int_s^{s+ds} e^{2\beta\tau} d\tau$$

This property of independent and Gaussian increment of $X^{u,\nu}(s)$ will be used in its simulation:

- Discretize the triangular plane $t <= v <= u <= T$ into equally spaced grid points and combine the points into a vector variable $X(s) =$ $(X^{u_1,u_1}(s), X^{u_2,u_1}(s), X^{u_2,u_2}(s), \ldots, X^{u_N,u_1}(s), \ldots, X^{u_N,u_N}(s))'$
- Initial value $X(0) = 0$;
- $X(s_{i+1}) - X(s_i)$ will be a vector of correlated Gaussian variables with Mean 0 and Variance-Covariance matrix Σ_{s_i};
- Repeat the last step, we can obtain $X(s_i) = \sum_{j=1}^{i}(X(s_{j+1}) - X(s_j))$ and further $g^{u,\nu}(s_i) = a(s_i, u, v) + b(s_i)X^{u,\nu}(s_i)$
- A sample from the discretized σ_t^2 distribution can be obtained as $2\sum_i \sum_{j=1}^{i} \sum_k \exp(g^{u_i,u_j}(s_k))c_1(u_i, u_j)(\Delta u)^2 \Delta s$
- Computation complexity will be $O(MN^4)$, where M is the number of grid points between 0 and t and N is the number of points between t and T

3.3.2 Pricing Errors

Pricing by Monte Carlo(MC) and log-normal approximation:

1. Simulate samples of σ_t^2 (here sample size n=10000) following last subsection;

2. Generate samples from log-normal distribution with the moments calculated from section 2.3;

3. Compute $C(0, \sigma_t^2)$ using σ_t^2 from the last two steps and then take the average separately to get two prices for the same contract; where

$$C(0,\sigma_t^2) = e^{-\int_0^{t+\Delta} f^u(0)du} \left\{ e^{\int_t^{t+\Delta} f^u(0)du} \Phi\left(\frac{\int_t^{t+\Delta} f^u(0)du - \Delta d}{\sigma_t} + \frac{\sigma_t}{2}\right) - e^{\Delta d} \Phi\left(\frac{\int_t^{t+\Delta} f^u(0)du - \Delta d}{\sigma_t} - \frac{\sigma_t}{2}\right) \right\}$$

4. Compare the two prices, report the relative pricing error for the setup;

5. Change parameter values and repeat the last four steps.

The price differences between MC method and L-N approximation $P_{MC}^D - P_{LN}$ can be decomposed as the sum of approximation error $P_{MC}^D - P_{LN}^D$ and discretization error from MC $P_{LN}^D - P_{LN}$. Here $P_{MC}^D = E[C(0, (\sigma_{MC}^D)^2)]$ is the option price calculated from the MC draws of σ_{MC}^D; $P_{LN}^D = E[C(0, (\sigma_{LN}^D)^2)]$ is the option price calculated from a L-N distributed variable $(\sigma_{LN}^D)^2$ with the same first two moments equal to those of the $(\sigma_{MC}^D)^2$, and $P_{LN} = E[C(0, \sigma_{LN}^2)]$ is the option price from a L-N distributed variable σ_{LN}^2 with first two moments calculated from last section.

Discretization Error from MC $P_{LN}^D - P_{LN}$

Since the first two moments of a L-N distributed variable $(\sigma_{LN}^D)^2$ converge to the corresponding moments of r.v. σ_{LN}^2, $(\sigma_{LN}^D)^2$ will converge to σ_{LN}^2 in distribution. Further because function $C(0, .)$ is bounded and continuous, $C(0, (\sigma_{LN}^D)^2)$ will converge to $C(0, \sigma_{LN}^2)$ in distribution as well. Then we have the result that Discretization Error $P_{LN}^D - P_{LN} = E[C(0, (\sigma_{LN}^D))^2] - E[C(0, \sigma_{LN}^2)]$ will converge to 0, as step size converges to 0.

Approximation Error $P^D_{MC} - P^D_{LN}$

Take Taylor's expansion around the first moment of σ_t^2 (denoted as $\overline{\sigma^2}$):

$$\begin{aligned} C(0) = \quad & C(0,\overline{\sigma^2}) + C^{(1)}(0,\overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2}) + \frac{1}{2}\ C^{(2)}(0,\overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2})^2 \\ & + \frac{1}{6}\ E\left[C^{(3)}(0,\xi)\ (\sigma_t^2 - \overline{\sigma^2})^3\right] \end{aligned} \tag{3.20}$$

Since $(\sigma^D_{MC})^2$ and $(\sigma^D_{LN})^2$ have the same first two moments, the first three terms in $P^D_{MC} = E[C(0,(\sigma^D_{MC})^2)]$ and $P^D_{LN} = E[C(0,(\sigma^D_{LN})^2)]$ are the same. Then we can conclude that approximation error $P^D_{MC} - P^D_{LN}$ will be purely from the difference between the last term. See Section 3.4 for the detailed formula for this term. Naturally, the approximation error from using log-normal instead of true distribution of σ_t^2 will be small, if either the distribution of σ_t^2 is close to log-normal or function $C(0,\sigma_t^2)$ is actually not far from quadratic form. Results from simulation study have shown that approximation error usually is very small in our setting.

Pricing Error from Simulation

We have actually carried out some simulations to check the above statement. It has been shown that, as we reduce step size, price differences between Monte Carlo and approximation have been gradually reduced to around 2% of the Cap price when computation complexity has prevented a further step size reduction. Of the 2% price difference (RMSE), only a very small portion (ApproxErr) is due to approximation error while the vast majority is due to discretization error. The result of the simulation study has also indicated that in order to have small enough discretization error from brute force Monte Carlo method, very fine grid points are necessary which may be

Table 3.1: Price differences from different correlation parameter k_i, where $c_i(u, v) = e^{-k_i|u-v|}$. Here, P_{LN} is the price from approximated LN distribution with matching mean and variance to σ_t^2; P_{MC}^D is the price from the Monte Carlo samples as σ_t^2; P_{LN}^D is the price from approximated LN distribution with the sample mean and variance matching the Monte Carlo samples of σ_t^2; $RMSE$ is the square-root of average squared difference between P_{LN} and P_{MC}^D; $TotalErr$ is the ratio of $RMSE$ and P_{MC}^D; $AppoxErr$ is the ratio between $P_{MC}^D - P_{LN}^D$ and P_{MC}^D. Other parameters values are chosen to be: $\alpha = -2, \beta = 1, \sigma_h = 1$; and the price of a at-the-money cap with $t = 1$ and $T = 1.25$ is considered and priced; procedures described in section 3.3.1 have been used to draw samples using Monte Carlo method, with sample size of 50,000; 200 grid points have been using between 0 and t, and 200 grids have been chosen between t and T.

k_1	k_2	P_{LN}	P_{MC}^D	P_{LN}^D	RMSE	TotalErr	AppoxErr
-8	-1	3.17	3.24	3.24	.076	2.35%	$< .1\%$
-4	-1	3.61	3.68	3.68	.076	2.07%	$< .1\%$
-2	-1	3.88	3.96	3.96	.081	2.05%	$< .1\%$
-1	-1	4.03	4.12	4.12	.087	2.11%	$< .1\%$
-0.5	-1	4.12	4.20	4.20	.083	1.98%	.1%
-0.25	-1	4.16	4.25	4.25	.086	2.02%	$< .1\%$
-0.125	-1	4.18	4.27	4.27	.086	2.01%	$< .1\%$
0	-1	4.21	4.29	4.29	.087	2.02%	.1%
-1	-8	4.02	4.10	4.10	.083	2.02%	$< .1\%$
-1	-4	4.03	4.11	4.11	.083	2.01%	$< .1\%$
-1	-2	4.03	4.12	4.12	.084	2.03%	$< .1\%$
-1	-1	4.03	4.12	4.12	.087	2.11%	$< .1\%$
-1	-0.5	4.04	4.12	4.12	.082	1.99%	.1%
-1	-0.125	4.04	4.12	4.12	.087	2.11%	$< .1\%$
-1	0	4.04	4.12	4.12	.084	2.04%	.2%

beyond even the modern computational power, while we can bypass the discretization step by approximating the distribution of the summary statistic from its first couple of moments.

Effect of Correlation Parameters

It can be seen from Table 1 that, the larger the correlation among forward rates (as controlled by k_1) the larger the Cap price will be, given other conditions the same. The rationale for this might be explained from the formula of first moment of σ_t. Where a smaller k_1 will make the integrand smaller. A possible economical explanation will be that when all shocks to rates of different maturities are highly correlated, the average rate between t and T should be more volatile thus the option should be more valuable. Similar explanation applies to the correlation (as controlled by k_2) for the volatility field. However, it can be seen that option price is sensitive to the value of k_1 while being very insensitive to the value of k_2. It shows that a more careful modeling for the correlation among shocks to forward rates might be worth pursuing, while we may model the correlation among volatilities less accurately without losing much.

3.4 Why the Approximation Works?

Consider the 1D case with no leverage effect, so our proposed approximation scheme focuses on the distribution of σ_t^2.

Recall (3.13), the cap price at time $t = 0$:

$$C(0) = E[C(0, \sigma_t^2)],$$

where the expectation is taken over σ_t^2 under the risk neutral measure and can be computed via Monte Carlo numerical integration. Let $\overline{\sigma^2} = E\sigma_t^2$ and $C^{(k)}(0, v) = \frac{\partial^k C(0,\sigma_t^2)}{\partial(\sigma_t^2)^k}\Big|_{\sigma_t^2 = v}$. We have the Taylor expansion around $\overline{\sigma^2}$:

$$\begin{aligned} C(0) &= C(0, \overline{\sigma^2}) + C^{(1)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2}) + \frac{1}{2}\ C^{(2)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2})^2 \\ &\quad + \frac{1}{6}\ E\left[C^{(3)}(0, \xi)\ (\sigma_t^2 - \overline{\sigma^2})^3\right] \end{aligned} \tag{3.21}$$

where ξ is between σ_t^2 and $\overline{\sigma^2}$. Notice the second term in (3.21) vanishes, and the leading term $C(0) = C(0, \overline{\sigma^2})$ is just the pricing formula in Kennedy (1994). Here is our extension of Kennedy's result: first, a second-order correction term $\frac{1}{2}\, C^{(2)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2})^2$ is added; however, we go beyond the approximation

$$C(0) \approx C(0, \overline{\sigma^2}) + \frac{1}{2}\ C^{(2)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2})^2$$

by incorporating some skewness corrections in the term $\frac{1}{6}\ E\left[C^{(3)}(0, \xi)\ (\sigma_t^2 - \overline{\sigma^2})^3\right]$. Although we do not include this term explicitly, skewness is taken into consideration when we choose a log-normal density instead of a normal density.

Let $\phi(x) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}$ be the standard normal density. In what follows, we attach the superscript (k) to various functions to denote their kth order partial derivatives. Recall

$$\begin{aligned}
\phi^{(1)}(x) &= \phi(x)(-x), \\
\phi^{(2)}(x) &= \phi(x)(x^2 - 1), \\
\phi^{(3)}(x) &= \phi(x)(-x^3 + 3x).
\end{aligned}$$

We have

$$C(0, \sigma_t^2) = \exp\left[-\int_0^{t+\Delta} f(0, u)\ du\right] \cdot A(\sigma_t^2) \tag{3.22}$$

where

$$A(\sigma_t^2) = B_t\ \Phi(b(\sigma_t^2) + \sqrt{\sigma_t^2}/2) - \ e^{\Delta K}\ \Phi(b(\sigma_t^2) - \sqrt{\sigma_t^2}/2) \tag{3.23}$$

with $B_t = \exp\left[\int_t^{t+\Delta} f(0, u)\ du\right]$ and $b(\sigma_t^2) = \left[\int_t^{t+\Delta} f(0, u)\ du - \Delta K\right]\ (\sigma_t^2)^{-1/2}$.

Note that

$$\begin{aligned}
b^{(1)}(\sigma_t^2) &= \frac{-1}{2}\ b(\sigma_t^2)\ (\sigma_t^2)^{-1}, \\
b^{(2)}(\sigma_t^2) &= \frac{3}{4}\ b(\sigma_t^2)\ (\sigma_t^2)^{-2}, \\
b^{(3)}(\sigma_t^2) &= \frac{-15}{8}\ b(\sigma_t^2)\ (\sigma_t^2)^{-3}.
\end{aligned}$$

Hence

$$
\begin{aligned}
A^{(1)}(\sigma_t^2) &= B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ [b^{(1)}(\sigma_t^2)+(\sigma_t^2)^{-1/2}/4] \\
&\quad -\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ [b^{(1)}(\sigma_t^2)-(\sigma_t^2)^{-1/2}/4] \\
&= B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ \left[\frac{-1}{2}\ b(\sigma_t^2)\ (\sigma_t^2)^{-1}+(\sigma_t^2)^{-1/2}/4\right] \\
&\quad -\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ \left[\frac{-1}{2}\ b(\sigma_t^2)\ (\sigma_t^2)^{-1}-(\sigma_t^2)^{-1/2}/4\right];
\end{aligned}
$$

and

$$
\begin{aligned}
A^{(2)}(\sigma_t^2) &= B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ \left[\frac{-1}{2}\ b(\sigma_t^2)\ (\sigma_t^2)^{-1}+\frac{1}{4}\ (\sigma_t^2)^{-1/2}\right]^2 \\
&\quad +B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ \left[\frac{3}{4}\ b(\sigma_t^2)\ (\sigma_t^2)^{-2}-\frac{1}{8}\ (\sigma_t^2)^{-3/2}\right] \\
&\quad -\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ \left[\frac{-1}{2}\ b(\sigma_t^2)\ (\sigma_t^2)^{-1}-\frac{1}{4}\ (\sigma_t^2)^{-1/2}\right]^2 \\
&\quad -\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ \left[\frac{3}{4}\ b(\sigma_t^2)\ (\sigma_t^2)^{-2}+\frac{1}{8}\ (\sigma_t^2)^{-3/2}\right] \\
&= B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ \left[\frac{b^2(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{16\sigma_t^2}-\frac{b(\sigma_t^2)}{4(\sigma_t^2)^{3/2}}+\frac{3b(\sigma_t^2)}{4(\sigma_t^2)^2}-\frac{1}{8(\sigma_t^2)^{3/2}}\right] \\
&\quad -\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ \left[\frac{b^2(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{16\sigma_t^2}+\frac{b(\sigma_t^2)}{4(\sigma_t^2)^{3/2}}+\frac{3b(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{8(\sigma_t^2)^{3/2}}\right] \\
&\triangleq B_t\ \phi(b(\sigma_t^2)+\sqrt{\sigma_t^2}/2)\ D_1(\sigma_t^2)-\ e^{\Delta K}\ \phi(b(\sigma_t^2)-\sqrt{\sigma_t^2}/2)\ D_2(\sigma_t^2),
\end{aligned}
$$

where

$$
\begin{aligned}
D_1(\sigma_t^2) &= \frac{b^2(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{16\sigma_t^2}-\frac{b(\sigma_t^2)}{4(\sigma_t^2)^{3/2}}+\frac{3b(\sigma_t^2)}{4(\sigma_t^2)^2}-\frac{1}{8(\sigma_t^2)^{3/2}}, \\
D_2(\sigma_t^2) &= \frac{b^2(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{16\sigma_t^2}+\frac{b(\sigma_t^2)}{4(\sigma_t^2)^{3/2}}+\frac{3b(\sigma_t^2)}{4(\sigma_t^2)^2}+\frac{1}{8(\sigma_t^2)^{3/2}}.
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
D_1^{(1)}(\sigma_t^2) &= \frac{-3b^2(\sigma_t^2)}{4(\sigma_t^2)^3} - \frac{1}{16(\sigma_t^2)^2} + \frac{b(\sigma_t^2)}{2(\sigma_t^2)^{5/2}} - \frac{15b(\sigma_t^2)}{8(\sigma_t^2)^3} + \frac{3}{16(\sigma_t^2)^{5/2}}, \\
D_2^{(1)}(\sigma_t^2) &= \frac{-3b^2(\sigma_t^2)}{4(\sigma_t^2)^3} - \frac{1}{16(\sigma_t^2)^2} - \frac{b(\sigma_t^2)}{2(\sigma_t^2)^{5/2}} - \frac{15b(\sigma_t^2)}{8(\sigma_t^2)^3} - \frac{3}{16(\sigma_t^2)^{5/2}}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
A^{(3)}(\sigma_t^2) &= B_t\ \phi(b(\sigma_t^2) + \sqrt{\sigma_t^2}/2)\ \left\{ \left[\frac{-b(\sigma_t^2)}{2\sigma_t^2} + \frac{1}{4(\sigma_t^2)^{1/2}} \right] D_1(\sigma_t^2) + D_1^{(1)}(\sigma_t^2) \right\} \\
&- e^{\Delta K}\ \phi(b(\sigma_t^2) - \sqrt{\sigma_t^2}/2)\ \left\{ \left[\frac{-b(\sigma_t^2)}{2\sigma_t^2} - \frac{1}{4(\sigma_t^2)^{1/2}} \right] D_2(\sigma_t^2) + D_2^{(1)}(\sigma_t^2) \right\}.
\end{aligned}
$$

The proposed approximation scheme yields a cap price that has a similar expression to (3.21) in which the symbol "$\sim$" is attached to those quantities with approximations involved. Here is a summary of our assessment on the approximation errors:

(i) In the formula

$$
\begin{aligned}
\widetilde{C}(0) &= C(0, \overline{\sigma^2}) + C^{(1)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2}) + \frac{1}{2}\ C^{(2)}(0, \overline{\sigma^2})\ E(\sigma_t^2 - \overline{\sigma^2})^2 \\
&+ \frac{1}{6}\ \widetilde{E}\left[C^{(3)}(0, \xi)\ (\sigma_t^2 - \overline{\sigma^2})^3 \right],
\end{aligned} \tag{3.24}
$$

the first three terms match their counterparts in (3.21) exactly based on our calculated moments $E\sigma_t^2$ and $E(\sigma_t^2)^2$.

(ii) To evaluate the approximation error in the last term, write

$$
E\left[C^{(3)}(0, \xi)\ (\sigma_t^2 - \overline{\sigma^2})^3 \right] = E_1 + E_2 + E_3 \tag{3.25}
$$

where $E_i = E\left[C^{(3)}(0, \xi)\ (\sigma_t^2 - \overline{\sigma^2})^3\ I_{A_i} \right]$, $i = 1, 2, 3$ and $A_1 = \{\xi > \overline{\sigma^2} + M\}$,

$A_2 = \{\overline{\sigma^2} \leq \xi \leq \overline{\sigma^2} + M\}$, $A_3 = \{\xi < \overline{\sigma^2}\}$ with a sufficiently large $M > 0$. Similarly, we can write

$$\widetilde{E}\left[C^{(3)}(0,\xi)\ (\sigma_t^2 - \overline{\sigma^2})^3\right] = \widetilde{E}_1 + \widetilde{E}_2 + \widetilde{E}_3. \tag{3.26}$$

(iii) Our calculation shows clearly that $C^{(3)}(0, v)$ drops to zero rapidly as $v \to \infty$ when holding other variables [t, Δ, K and the forward rate $f(\cdot)$] fixed. Therefore, the error $|\widetilde{E}_1 - E_1|$ should be small.

(iv) Note that for $i = 2, 3$,

$$|\widetilde{E}_i - E_i| \leq \sup\{|C^{(3)}(0,\xi)| : \ \xi \in E_i\} \ \left[|\widetilde{E}(|\sigma_t^2 - \overline{\sigma^2}|^3 I_{A_2}) - E(|\sigma_t^2 - \overline{\sigma^2}|^3 I_{A_2})|\right],$$

in which the skewness factor will affect the result of approximation. That explains why using a log-normal distribution for σ_t^2 would be better than using a normal distribution.

Chapter 4

Option Pricing With Leverage Effect

While some previous papers argue that innovations in interest rate levels are largely uncorrelated with innovations in the volatility of interest rates (e.g., Ball and Torous (1999), Chen and Scott (2001) and Heidari and Wu (2003)), Trolle and Schwartz (2009) shows evidence that the correlation between forward rates and their volatilities is important in capturing the implied volatility skewness observable in the option market. In this chapter, we would like to extend the framework described in the last chapter to this more realistic case. In other words, we would like the innovations to forward rates and their volatilities to be correlated. The idea will be the same as in the previous chapter; however both pricing formulas and computation will be a lot messier, as there will be several correction terms entering into the pricing formula to account for the correlation between the two processes.

The structure of this chapter is similar to the previous one. First we derive the pricing formula conditioning on the volatility process, with the additional assumption that the two random noises processes for forward rates and their volatilities are correlated.

Note that there are now three summary statistics presented in the formula involving a total of seven variables related to the volatility process. In the second section, we derive the formulas to calculate the first two moments of this seven-variate vector, through which a trivariate Gaussian vector is proposed to approximate the summary statistics. Lastly, we test the performance of this Gaussian approximation in the calculation of option prices.

4.1 Option Pricing Formula

Proposition 2. *Under the same setting of last chapter, if instead of independent innovations, we consider correlated innovations to take into account of leverage effect.*

$$\begin{aligned} d\mathbf{f}(t,T) &= \mu(t,T)\ dt + \sigma(t,T)\ \left[\sqrt{1-\rho^2}\ d\mathbf{W}_1(t,T) + \rho\ d\mathbf{W}_2(t,T)\right], \quad (4.1) \\ dlog\sigma^2(t,T) &= [\alpha - \beta\ log\sigma^2(t,T)]\ dt + \sigma_h\ d\mathbf{W}_2(t,T) \end{aligned}$$

If we denote $\sigma(s,T) := \exp[h(s,T)/2]$*, the drift* $\mu(t,T)$ *satisfies the no-arbitrage condition:*

$$\mu(s,T) = \sigma(s,T)\int_0^T \sigma(s,u)\left[\sqrt{1-\rho^2}c_1(T,u) + \rho c_2(T,u)\right] du.$$

Then conditioning on both the past and the future of volatility structure, the time 0 price of the cap will be given by :

$$\begin{aligned} e^{-\int_0^{t+\Delta} f(0,u)du} \Bigg\{ & e^{\int_t^{t+\Delta} f(0,u)du} e^{-Z_1} \Phi\Big(\frac{\int_t^{t+\Delta} f(0,u)du - \Delta K + Z_2 - Z_1}{\sqrt{Z_3}} + \frac{\sqrt{Z_3}}{2}\Big) \\ & - e^{\Delta K} e^{-Z_2} \Phi\Big(\frac{\int_t^{t+\Delta} f(0,u)du - \Delta K + Z_2 - Z_1}{\sqrt{Z_3}} - \frac{\sqrt{Z_3}}{2}\Big) \Bigg\} \quad (4.2) \end{aligned}$$

where $Z_1 = \rho(Y_1 + Y_2) + X_1 + X_2$, $Z_2 = \rho Y_3 + X_3$ *and* $Z_3 = 2(1-\rho^2)X_4$ *as defined in (4.3) – (4.5).*

Proof: Denote as before $N_1 = \Delta f^{\Delta}(t,t) = \int_t^{t+\Delta} f(t,u)du$ and $N_2 = \int_0^{t+\Delta} R_u du = \int_0^{t+\Delta} f(u,u)du$, then the time 0 price of a cap with strike price K between the period $[t, t+\Delta]$ from expectation (3.2) is given by (3.5):

$$\begin{aligned} E\left(e^{-N_2}(e^{N_1} - e^{\Delta K})^+\right) &= e^{E(N_1-N_2)+\frac{Var(N_1-N_2)}{2}} \Phi\left(\frac{E(N_1) + Var(N_1) - Cov(N_1,N_2) - \Delta K}{\sqrt{Var(N_1)}}\right) \\ &- e^{\Delta K - E(N_2) + \frac{Var(N_2)}{2}} \Phi\left(\frac{E(N_1) - Cov(N_1,N_2) - \Delta K}{\sqrt{Var(N_1)}}\right) \end{aligned}$$

Since

$$\begin{aligned} N1 = \int_t^{t+\Delta} f(0,u)du + \int_t^{t+\Delta}\int_0^t \mu(u,v)dudv + \rho \int_t^{t+\Delta}\int_0^t \sigma(u,v)d\mathbf{W}_2(u,v)dv \\ + \sqrt{1-\rho^2}\int_t^{t+\Delta}\int_0^t \sigma(u,v)d\mathbf{W}_1(u,v)dv \end{aligned}$$

$$\begin{aligned} N2 = \int_0^{t+\Delta} f(0,u)du + \int_0^{t+\Delta}\int_0^v \mu(u,v)dudv + \rho \int_0^{t+\Delta}\int_0^v \sigma(u,v)d\mathbf{W}_2(u,v)dv \\ + \sqrt{1-\rho^2}\int_0^{t+\Delta}\int_0^v \sigma(u,v)d\mathbf{W}_1(u,v)dv \end{aligned}$$

and (N_1, N_2) is a bivariate normal vector. Thus

$$\begin{aligned} -N1 + N2 = \int_0^t f(0,u)du + \int_0^t\int_0^v \mu(u,v)dudv + \rho\int_0^t\int_0^v \sigma(u,v)d\mathbf{W}_2(u,v)dv \\ + \int_t^{t+\Delta}\int_t^v \mu(u,v)dudv + \rho\int_t^{t+\Delta}\int_t^v \sigma(u,v)d\mathbf{W}_2(u,v)dv \\ + \sqrt{1-\rho^2}\int_0^t\int_0^v \sigma(u,v)d_u\mathbf{W}_1(u,v)dv + \sqrt{1-\rho^2}\int_t^{t+\Delta}\int_t^v \sigma(u,v)d\mathbf{W}_1(u,v)dv \end{aligned}$$

Further as

$$
\begin{aligned}
\int_0^t \int_0^v \mu(u,v)dudv = & \int_0^t \int_0^{v_1} \sigma(u,v_1) \int_u^{v_1} \sigma(u,v_2)\left[\sqrt{1-\rho^2}c_1(v_1,v_2)+\rho c_2(v_1,v_2)\right] dv_2 dudv_1 \\
= & \int_0^t \int_0^{v_2} \int_0^{v_1} \sigma(u,v_1)\sigma(u,v_2)\left[\sqrt{1-\rho^2}c_1(v_1,v_2)+\rho c_2(v_1,v_2)\right] dudv_1dv_2
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\int_0^t \int_0^v \sigma(u,v)dW_1(u,v)dv\right) = & \int_0^t \int_0^t \int_0^{v_1\wedge v_2} \sigma(u,v_1)\sigma(u,v_2)c_1(v_1,v_2)dudv_1dv_2 \\
= & \; 2\int_0^t \int_0^{v_2} \int_0^{v_1} \sigma(u,v_1)\sigma(u,v_2)c_1(v_1,v_2)dudv_1dv_2
\end{aligned}
$$

We can have the following quantities to be substituted into equation (3.5) .

$$
\begin{aligned}
& -E(N_1-N_2)-\frac{Var(N_1-N_2)}{2} \\
= & \int_0^t f(0,u)du+\rho\int_0^t \int_0^v \sigma(u,v)d\mathbf{W}_2(u,v)dv+\rho\int_t^{t+\Delta}\int_t^v \sigma(u,v)d\mathbf{W}_2(u,v)dv \\
+ & \int_0^t \int_0^{v_2} \int_0^{v_1} \sigma(u,v_1)\sigma(u,v_2)\left[\sqrt{1-\rho^2}(1-\sqrt{1-\rho^2})c_1(v_1,v_2)+\rho c_2(v_1,v_2)\right] dudv_1dv_2 \\
+ & \int_t^{t+\Delta} \int_t^{v_2} \int_t^{v_1} \sigma(u,v_1)\sigma(u,v_2)\left[\sqrt{1-\rho^2}(1-\sqrt{1-\rho^2})c_1(v_1,v_2)+\rho c_2(v_1,v_2)\right] dudv_1dv_2 \\
:= & \int_0^t f(0,u)du+\rho Y_1+\rho Y_2+X_1+X_2 := \int_0^t f(0,u)du+Z_1 \qquad (4.3)
\end{aligned}
$$

$$
\begin{aligned}
& E(N_2)-\frac{Var(N_2)}{2} \\
= & \int_0^{t+\Delta} f(0,u)du+\rho\int_0^{t+\Delta}\int_0^v \sigma(u,v)d\mathbf{W}_2(u,v)dv \\
& +\int_0^{t+\Delta} \int_0^{v_2} \int_0^{v_1} \sigma(u,v_1)\sigma(u,v_2)\left[\sqrt{1-\rho^2}(1-\sqrt{1-\rho^2})c_1(v_1,v_2)+\rho c_2(v_1,v_2)\right] dudv_1dv_2 \\
:= & \int_0^{t+\Delta} f(0,u)du+\rho Y_3+X_3 := \int_0^{t+\Delta} f(0,u)du+Z_2 \qquad (4.4)
\end{aligned}
$$

$$
\begin{aligned}
Var(N_1) &= (1-\rho^2)\int_t^{t+\Delta}\int_t^{t+\Delta}\int_0^t \sigma(u,v_1)\sigma(u,v_2)c_1(v_1,v_2)dudv_1dv_2 \\
&= 2(1-\rho^2)\int_t^{t+\Delta}\int_t^{v_2}\int_0^t \sigma(u,v_1)\sigma(u,v_2)c_1(v_1,v_2)dudv_1dv_2 \\
&:= 2(1-\rho^2)X_4 := Z_3 \qquad (4.5)
\end{aligned}
$$

and

$$
E(N_1)+\frac{Var(N_1)}{2}-Cov(N_1,N_2)=E(N_2)-\frac{Var(N_2)}{2}+E(N_1-N_2)+\frac{Var(N_1-N_2)}{2} \qquad (4.6)
$$

Substitute into the formula we obtain:

$$
\begin{aligned}
E\left(e^{-N_2}(e^{N_1}-e^{\Delta K})^+\right) &= e^{E(N_1-N_2)+\frac{Var(N_1-N_2)}{2}}\Phi(\frac{E(N_1)+Var(N_1)-Cov(N_1,N_2)-\Delta K}{\sqrt{Var(N_1)}}) \\
&\quad -e^{\Delta K-E(N_2)+\frac{Var(N_2)}{2}}\Phi(\frac{E(N_1)-Cov(N_1,N_2)-\Delta K}{\sqrt{Var(N_1)}}) \\
&= e^{-\int_0^{t+\Delta}f(0,u)du}\left\{e^{\int_t^{t+\Delta}f(0,u)du}e^{-Z_1}\Phi(\frac{\int_t^{t+\Delta}f(0,u)du-\Delta K+Z_2-Z_1}{\sqrt{Z_3}}+\frac{\sqrt{Z_3}}{2})\right. \\
&\quad \left. -e^{\Delta K}e^{-Z_2}\Phi(\frac{\int_t^{t+\Delta}f(0,u)du-\Delta K+Z_2-Z_1}{\sqrt{Z_3}}-\frac{\sqrt{Z_3}}{2})\right\}
\end{aligned}
$$

QED.

4.2 Moments of Characterizing Variables

Similar to the case when shocks to forward rates and their volatility are independent, in order to get the price of the cap we will need to find the joint distribution of the summary statistics $(Z_1, Z_2, Z_3)^T$. Obviously it is very difficult to find the exact distribution for this three-dimension vector, and thus we will try to find its best proxy within a multi-variate parametric family. The easiest candidate is the Gaussian family. But considering Z_3 is an integrated volatility and alway positive, we will use a jointly Gaussian distribution to approximate vector $(Z_1, Z_2, ln(Z_3))^T$, and use it as the summary statistics instead. We will find the proxy within Gaussian family through the method of moments.

Fact 1: For two jointly normal random variables N_1 and N_2, the following result holds

$$Cov\left(e^{N_1},\ N_2\right) = Cov(N_1, N_2)\ e^{EN_1 + Var(N_1)/2} \tag{4.7}$$

Under the Gaussian assumption for $(Z_1, Z_2, ln(Z_3))^T$ and using **Fact 1**, its mean and variance-covariance matrix can be calculated from the corresponding values for $(Z_1, Z_2, Z_3)^T$. The mean and variance-covariance matrix for $(Z_1, Z_2, Z_3)^T$ can be found through the length-7 vector $(X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3)^T$ whose first two moments can be calculated based on the following formulas.

Denote $c^*(v_1, v_2) = \sqrt{1-\rho^2}(1-\sqrt{1-\rho^2})c_1(v_1, v_2) + \rho c_2(v_1, v_2)$, then

$$\begin{aligned}
EX_1 &= \int_0^t \int_0^{v_2} \int_0^{v_1} e^{H_1(u,v_1,v_2)} c^*(v_1, v_2) du dv_1 dv_2 && (4.8)\\
EX_2 &= \int_t^{t+\Delta} \int_t^{v_2} \int_t^{v_1} e^{H_1(u,v_1,v_2)} c^*(v_1, v_2) du dv_1 dv_2 && (4.9)\\
EX_3 &= \int_0^{t+\Delta} \int_0^{v_2} \int_0^{v_1} e^{H_1(u,v_1,v_2)} c^*(v_1, v_2) du dv_1 dv_2 && (4.10)\\
EX_4 &= \int_t^{t+\Delta} \int_t^{v_2} \int_0^{t} e^{H_1(u,v_1,v_2)} c_1(v_1, v_2) du dv_1 dv_2 && (4.11)\\
EY_i &= 0,\ \forall i = 1, 2, 3. && (4.12)
\end{aligned}$$

$$
\begin{aligned}
Var(Y_1) &= E\int_0^t\int_0^t\int_0^{v_1\wedge v_2}\sigma(u,v_1)\sigma(u,v_2)c_2(v_1,v_2)dudv_1dv_2 \\
&= 2\ E\int_0^t\int_0^{v_2}\int_0^{v_1}\sigma(u,v_1)\sigma(u,v_2)c_2(v_1,v_2)dudv_1dv_2 \\
&= 2\ \int_0^t\int_0^{v_2}\int_0^{v_1}e^{H_1(u,v_1,v_2)}c_2(v_1,v_2)dudv_1dv_2 \qquad (4.13)\\
Var(Y_2) &= 2\ \int_t^{t+\Delta}\int_t^{v_2}\int_t^{v_1}e^{H_1(u,v_1,v_2)}c_2(v_1,v_2)dudv_1dv_2 \qquad (4.14)\\
Var(Y_3) &= 2\ \int_0^{t+\Delta}\int_0^{v_2}\int_0^{v_1}e^{H_1(u,v_1,v_2)}c_2(v_1,v_2)dudv_1dv_2 \qquad (4.15)\\
Cov(Y_1,Y_2) &= 0 \qquad (4.16)\\
Cov(Y_1,Y_3) &= \int_0^{t+\Delta}\int_0^t\int_0^{v_1\wedge v_2}e^{H_1(u,v_1,v_2)}c_2(v_1,v_2)dudv_1dv_2 \qquad (4.17)\\
Cov(Y_2,Y_3) &= \int_0^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_1\wedge v_2}e^{H_1(u,v_1,v_2)}c_2(v_1,v_2)dudv_1dv_2 \qquad (4.18)
\end{aligned}
$$

$$
\begin{aligned}
EX_1^2 &= \int_0^t\int_0^t\int_0^{v_{22}}\int_0^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_2^2 &= \int_t^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_{22}}\int_t^{v_{21}}\int_t^{v_{12}}\int_t^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_3^2 &= \int_0^{t+\Delta}\int_0^{t+\Delta}\int_0^{v_{22}}\int_0^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_4^2 &= \int_t^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_{22}}\int_t^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c_1(v_{11},v_{21})c_1(v_{12},v_{22})\ d\mathbf{u}\\
EX_1X_2 &= \int_t^{t+\Delta}\int_0^t\int_t^{v_{22}}\int_0^{v_{21}}\int_t^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_1X_3 &= \int_0^{t+\Delta}\int_0^t\int_0^{v_{22}}\int_0^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_1X_4 &= \int_t^{t+\Delta}\int_0^t\int_t^{v_{22}}\int_0^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c_1(v_{12},v_{22})\ d\mathbf{u}\\
EX_2X_3 &= \int_0^{t+\Delta}\int_t^{t+\Delta}\int_0^{v_{22}}\int_t^{v_{21}}\int_0^{v_{12}}\int_t^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c^*(v_{12},v_{22})\ d\mathbf{u}\\
EX_2X_4 &= \int_t^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_{22}}\int_t^{v_{21}}\int_0^{v_{12}}\int_t^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c_1(v_{12},v_{22})\ d\mathbf{u}\\
EX_3X_4 &= \int_t^{t+\Delta}\int_0^{t+\Delta}\int_t^{v_{22}}\int_0^{v_{21}}\int_0^{v_{12}}\int_0^{v_{11}}e^{H_2(u_1,u_2,v_{11},v_{12},v_{21},v_{22})}c^*(v_{11},v_{21})c_1(v_{12},v_{22})\ d\mathbf{u}
\end{aligned}
$$

where we abbreviate $d\mathbf{u} := du_1 du_2 dv_{11} dv_{12} dv_{21} dv_{22}$ in EX_1^2 — EX_3X_4.

To calculate $Cov(X_i, Y_j)$, we will make use of **Fact 1** and also the fact that when $0 < u_3 < u$,

$$\begin{aligned} h(u, v_1)\ dW_2(u_3, v_3) &= \int_0^u e^{-\int_\tau^u \beta(s,v_1)ds}\sigma_h(\tau, v_1) dW_2(\tau, v_1)\ dW_2(u_3, v_3) \\ &= e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3, v_1) c_2(v_1, v_3)\ du_3\ I_{\{u_3<u\}} \end{aligned}$$

Now we can calculate the following result:

$$\begin{aligned} & Cov(X_1, Y_1) \\ = \ & E\left[\int_0^t \int_0^{v_2} \int_0^{v_1} \sigma(u, v_1)\sigma(u, v_2) c^*(v_1, v_2) du dv_1 dv_2\right] \left[\int_0^t \int_0^v \sigma(u_3, v_3) d\mathbf{W}_2(u_3, v_3) dv_3\right] \\ = \ & E\int_0^t \int_0^t \int_0^{v_2} \int_0^{v_1} \int_0^{v_3} \sigma(u, v_1)\sigma(u, v_2)\sigma(u_3, v_3) c^*(v_1, v_2) dW_2(u_3, v_3) du dv_1 dv_2 dv_3 \\ = \ & \int_0^t \int_0^t \int_0^{v_2} \int_0^{v_1} c^*(v_1, v_2) E\int_0^{v_3} \sigma(u, v_1)\sigma(u, v_2)\sigma(u_3, v_3) dW_2(u_3, v_3) du dv_1 dv_2 dv_3 \\ = \ & \frac{1}{2}\int_0^t \int_0^t \int_0^{v_2} \int_0^{v_1} c^*(v_1, v_2) \int_0^{u\wedge v_3} [e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3, v_1) c_2(v_1, v_3) \\ & \quad + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3, v_2) c_2(v_2, v_3)] e^{H_3(u_3,u,v_1,v_2,v_3)}\ du_3 du dv_1 dv_2 dv_3 \end{aligned} \tag{4.19}$$

where $du_3 du dv_1 dv_2 dv_3$ is abbreviated as $d\mathbf{v}$ and

$$
\begin{aligned}
H_3(u_3, u, v_1, v_2, v_3) &= E(log\sigma(u, v_1) + log\sigma(u, v_2) + log\sigma(u_3, v_3)) \\
&+ \frac{Var(log\sigma(u, v_1) + log\sigma(u, v_2) + log\sigma(u_3, v_3))}{2} \\
&= \frac{1}{2}E(log\sigma^2(u, v_1) + log\sigma^2(u, v_2) + log\sigma^2(u_3, v_3)) \\
&+ \frac{1}{8}Var(log\sigma^2(u, v_1) + log\sigma^2(u, v_2) + log\sigma^2(u_3, v_3)) \\
&= \frac{1}{8}[\int_0^u e^{-\int_\tau^u 2\beta(s,v_1)ds}\sigma_h^2(\tau, v_1)d\tau + \int_0^u e^{-\int_\tau^u 2\beta(s,v_2)ds}\sigma_h^2(\tau, v_2)d\tau \\
&+ \int_0^{u_3} e^{-\int_\tau^{u_3} 2\beta(s,v_3)ds}\sigma_h^2(\tau, v_3)d\tau \\
&+ 2\int_0^u e^{-\int_\tau^u(\beta(s,v_1)+\beta(s,v_2))ds}\sigma_h(\tau, v_1)\sigma_h(\tau, v_2)c(v_1, v_2)d\tau \\
&+ 2\int_0^{u\wedge u_3} e^{-\int_\tau^{u\wedge u_3}(\beta(s,v_1)+\beta(s,v_3))ds}\sigma_h(\tau, v_1)\sigma_h(\tau, v_3)c(v_1, v_3)d\tau \\
&+ 2\int_0^{u\wedge u_3} e^{-\int_\tau^{u\wedge u_3}(\beta(s,v_2)+\beta(s,v_3))ds}\sigma_h(\tau, v_2)\sigma_h(\tau, v_3)c(v_2, v_3)d\tau]
\end{aligned}
\tag{4.20}
$$

Similarly, we can obtain other cross products:

$$\begin{aligned}
& Cov(X_1, Y_2) \qquad (4.21)\\
= \; & \frac{1}{2}\int_t^{t+\Delta}\int_0^t\int_0^{v_2}\int_0^{v_1} c^*(v_1,v_2)\int_t^{u\wedge v_3}\left[c_2(v_1,v_3)+c_2(v_2,v_3)\right]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}\\
= \; & 0
\end{aligned}$$

$$\begin{aligned}
& Cov(X_1, Y_3) \qquad (4.22)\\
= \; & \frac{1}{2}\int_0^{t+\Delta}\int_0^t\int_0^{v_2}\int_0^{v_1} c^*(v_1,v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3,v_1)c_2(v_1,v_3)\\
& + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3,v_2)c_2(v_2,v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}
\end{aligned}$$

$$\begin{aligned}
& Cov(X_2, Y_1) \qquad (4.23)\\
= \; & \frac{1}{2}\int_0^{t}\int_t^{t+\Delta}\int_t^{v_2}\int_t^{v_1} c^*(v_1,v_2)\int_0^{u\wedge v_3}\left[c_2(v_1,v_3)+c_2(v_2,v_3)\right]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}\\
= \; & 0
\end{aligned}$$

$$\begin{aligned}
& Cov(X_2, Y_2) \qquad (4.24)\\
= \; & \frac{1}{2}\int_t^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_2}\int_t^{v_1} c^*(v_1,v_2)\int_t^{u\wedge v_3}[e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3,v_1)c_2(v_1,v_3)\\
& + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3,v_2)c_2(v_2,v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}
\end{aligned}$$

$$\begin{aligned}
& Cov(X_2, Y_3) \qquad (4.25)\\
= \; & \frac{1}{2}\int_0^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_2}\int_t^{v_1} c^*(v_1,v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3,v_1)c_2(v_1,v_3)\\
& + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3,v_2)c_2(v_2,v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}
\end{aligned}$$

$$\begin{aligned}
& Cov(X_3, Y_1) \qquad (4.26)\\
= \; & \frac{1}{2}\int_0^{t}\int_0^{t+\Delta}\int_0^{v_2}\int_0^{v_1} c^*(v_1,v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3,v_1)c_2(v_1,v_3)\\
& + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3,v_2)c_2(v_2,v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}
\end{aligned}$$

$$\begin{aligned}
& Cov(X_3, Y_2) \qquad (4.27)\\
= \; & \frac{1}{2}\int_t^{t+\Delta}\int_0^{t+\Delta}\int_0^{v_2}\int_0^{v_1} c^*(v_1,v_2)\int_t^{u\wedge v_3}[e^{-\int_{u_3}^u \beta(s,v_1)ds}\sigma_h(u_3,v_1)c_2(v_1,v_3)\\
& + e^{-\int_{u_3}^u \beta(s,v_2)ds}\sigma_h(u_3,v_2)c_2(v_2,v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\,d\mathbf{v}
\end{aligned}$$

$$(4.28)$$

$$
\begin{aligned}
& Cov(X_3, Y_3) \qquad (4.29)\\
= \ & \frac{1}{2}\int_0^{t+\Delta}\int_0^{t+\Delta}\int_0^{v_2}\int_0^{v_1} c^*(v_1, v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^{u}\beta(s,v_1)ds}\sigma_h(u_3, v_1)c_2(v_1, v_3)\\
& \quad + e^{-\int_{u_3}^{u}\beta(s,v_2)ds}\sigma_h(u_3, v_2)c_2(v_2, v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\ d\mathbf{v}
\end{aligned}
$$

$$
\begin{aligned}
& Cov(X_4, Y_1) \qquad (4.30)\\
= \ & \frac{1}{2}\int_0^{t}\int_t^{t+\Delta}\int_t^{v_2}\int_0^{v_1} c_1(v_1, v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^{u}\beta(s,v_1)ds}\sigma_h(u_3, v_1)c_2(v_1, v_3)\\
& \quad + e^{-\int_{u_3}^{u}\beta(s,v_2)ds}\sigma_h(u_3, v_2)c_2(v_2, v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\ d\mathbf{v}
\end{aligned}
$$

$$
\begin{aligned}
& Cov(X_4, Y_2) \qquad (4.31)\\
= \ & \frac{1}{2}\int_t^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_2}\int_0^{v_1} c_1(v_1, v_2)\int_t^{u\wedge v_3}[c_2(v_1, v_3) + c_2(v_2, v_3)]\, e^{H_3(u_3,u,v_1,v_2,v_3)}\ d\mathbf{v}\\
= \ & 0
\end{aligned}
$$

$$
\begin{aligned}
& Cov(X_4, Y_3) \qquad (4.32)\\
= \ & \frac{1}{2}\int_0^{t+\Delta}\int_t^{t+\Delta}\int_t^{v_2}\int_0^{v_1} c_1(v_1, v_2)\int_0^{u\wedge v_3}[e^{-\int_{u_3}^{u}\beta(s,v_1)ds}\sigma_h(u_3, v_1)c_2(v_1, v_3)\\
& \quad + e^{-\int_{u_3}^{u}\beta(s,v_2)ds}\sigma_h(u_3, v_2)c_2(v_2, v_3)]e^{H_3(u_3,u,v_1,v_2,v_3)}\ d\mathbf{v}
\end{aligned}
$$

With the above formulas, the first two moments for vector $Z = (\rho(Y_1 + Y_2) + X_1 + X_2, \rho Y_3 + X_3, 2(1 - \rho^2)X_4)^T$ can be calculated as having mean

$$(EX_1 + EX_2,\ EX_3,\ 2(1 - \rho^2)EX_4)^T \qquad (4.33)$$

and variance-covariance matrix $\Sigma = \{\Sigma_{ij}\}$ with

$$\begin{aligned}
\Sigma_{11} &= \rho^2 Var(Y_1) + \rho^2 Var(Y_2) + Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) \\
&\quad +2Cov(X_1, Y_1) + 2Cov(X_2, Y_2) \\
&= \rho^2 Var(Y_1) + \rho^2 Var(Y_2) + EX_1^2 + EX_2^2 + 2EX_1X_2 + (EX_1 - EX_2)^2 \\
&\quad +2Cov(X_1, Y_1) + 2Cov(X_2, Y_2) \qquad (4.34) \\
\Sigma_{12} &= \Sigma_{21} = \rho^2\ Cov(Y_1, Y_3) + \rho^2\ Cov(Y_2, Y_3) + Cov(X_1, X_3) + Cov(X_2, X_3) \\
&\quad +\rho Cov(X_1, Y_3) + \rho Cov(X_2, Y_3) + \rho Cov(X_3, Y_1) + \rho Cov(X_3, Y_2) \\
&= \rho^2 Cov(Y_1, Y_3) + \rho^2 Cov(Y_2, Y_3) + EX_1X_3 + EX_2X_3 - EX_1EX_3 - EX_2X_3 \\
&\quad +\rho Cov(X_1, Y_3) + \rho Cov(X_2, Y_3) + \rho Cov(X_3, Y_1) + \rho Cov(X_3, Y_2) \qquad (4.35) \\
\Sigma_{13} &= \Sigma_{31} \qquad (4.36) \\
&= 2(1-\rho^2)\ [Cov(X_1, X_4) +\ Cov(X_2, X_4) + \rho Cov(X_4, Y_1) + \rho Cov(X_4, Y_1)] \\
&= 2(1-\rho^2)\,[EX_1X_4 + EX_2X_4 - EX_1EX_4 - EX_2EX_4 + \rho EX_4Y_1 + \rho EX_4Y_1] \\
\Sigma_{22} &= \rho^2\ Var(Y_3) + Var(X_3) + 2\rho Cov(X_3, Y_3) \\
&= \rho^2\ Var(Y_3) + EX_3^2 - (EX_3)^2 + 2\rho Cov(X_3, Y_3) \qquad (4.37) \\
\Sigma_{23} &= \Sigma_{32} = 2(1-\rho^2)\ Cov(X_3, X_4) + 2\rho(1-\rho^2) Cov(X_4, Y_3) \\
&= 2(1-\rho^2)\ \ [EX_3X_4 - EX_3EX_4 + \rho Cov(X_4, Y_3)] \qquad (4.38) \\
\Sigma_{33} &= 4(1-\rho^2)^2\ Var(X_4) \\
&= 4(1-\rho^2)^2\ \left[EX_4^2 - (EX_4)^2\right]. \qquad (4.39)
\end{aligned}$$

4.3 Numerical Study

In this section, we will follow a similar structure as displayed in the previous chapter to study 1) the performance of the approximation scheme in option pricing under this more general framework; 2) the effect of model parameters over option price. In particular, we would like to study the effect of correlation parameter between the two shocks. To achieve 1), we will make use of results in the section 4.2, calculating the first two moments of the summary statistics, and then generate Gaussian vector from the calculated moments to obtain the approximated option price. Similar to last chapter, we will consider the simple case where the parameters are constant, but all parameters can be easily extended to be deterministic function of time.

4.3.1 Option Pricing

From the equation of first section, option price from model of this chapter can be obtained by first finding the joint distribution of the summary statistics $(Z_1, Z_2, Z_3)^T$, and then calculating the mean values for function, where $Z_1 = \rho Y_1 + \rho Y_2 + X_1 + X_2$, $Z_2 = \rho Y_3 + X_3$ and $Z_3 = 2(1-\rho^2)X_4$.

$$
\begin{aligned}
X_1 &= \int_0^t \int_0^{v_2} \int_0^{v_1} \sigma(u, v_1)\sigma(u, v_2)c^*(v_1, v_2)du dv_1 dv_2 \\
X_2 &= \int_t^{t+\Delta} \int_t^{v_2} \int_t^{v_1} \sigma(u, v_1)\sigma(u, v_2)c^*(v_1, v_2)du dv_1 dv_2 \\
X_3 &= \int_0^{t+\Delta} \int_0^{v_2} \int_0^{v_1} \sigma(u, v_1)\sigma(u, v_2)c^*(v_1, v_2)du dv_1 dv_2 \\
X_4 &= \int_t^{t+\Delta} \int_t^{v_2} \int_0^{t} \sigma(u, v_1)\sigma(u, v_2)c_1(v_1, v_2)du dv_1 dv_2 \\
Y_1 &= \int_0^t \int_0^{v} \sigma(u, v)d\mathbf{W}_2(u, v)dv \\
Y_2 &= \int_t^{t+\Delta} \int_t^{v} \sigma(u, v)d\mathbf{W}_2(u, v)dv \\
Y_3 &= \int_0^{t+\Delta} \int_0^{v} \sigma(u, v)d\mathbf{W}_2(u, v)dv
\end{aligned}
$$

where $c^*(v_1, v_2) = \sqrt{1-\rho^2}(1-\sqrt{1-\rho^2})c_1(v_1, v_2) + \rho c_2(v_1, v_2)$.

Method 1:

Since $log\sigma^2(u,v) = e^{-\beta u}\ log\sigma^2(0,v) + \frac{\alpha}{\beta}(1-e^{-\beta u}) + \sigma e^{-\beta u}\int_0^u e^{\beta s}d\mathbf{W}_2(s,v),$ all the summary statistics will be solely determined by the history of $W_2(u,v),\ 0<u<v<T$. The simulated option price can be obtained through the following process:

1. Simulate discretized triangular plane $W_2(u,v),\ 0<u<v<T$
2. calculate $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3$ and thus Z_1, Z_2, Z_3
3. calculate option price $C_0(K, Z_1, Z_2, Z_3)$ where K is the strike rate
4. repeat the above process to get a series of simulated option payout under the martingale measure, and then take the average to get the option price as of time 0;

Method 2:

We also approximate the option price by

1. Calculate first two moments of the summary statistics vector (Z_1, Z_2, Z_3),
2. Assume $(Z_1, Z_2, log(Z_3))^T$ is Gaussian, and make use of **Fact 1** to calculate its mean and variance-covariance matrix,
3. generate normal vector with matching first two moments to $(Z_1, Z_2, log(Z_3))^T$,
4. calculate option price from the simulated normal vector

Now we can compare the option price from the two different methods.

4.3.2 Pricing Errors

The brute force MC of **Method 1** will introduce discretization error, which theoretically could be reduced to 0 if we can increase the number of grid points to infinite and generate even more samples. But practically speaking, this is impossible and we shall find that it is actually very difficult to reduce due to the computation power it requires for the random filed process.

The Gaussian approximation of the summary statistics of **Method 2** will introduce some approximation Error. It may be true that the joint distribution of the summary statistics are not Gaussian, but under certain condition similar to last chapter, its effect on the pricing can be bounded and small.

The following table 4.1 will show that option price from **Method 1** are actually convergent to **Method 2** when we increase the number of grid points; and furthermore it also illustrates that option prices from the two methods are close for a number of different combination of parameter settings.

4.3.3 Effect of Correlation Parameters

Correlation Parameter k_1 and k_2

It has been seen in the previous chapter that correlation among forward rates has a large effect on option price while correlation among their volatilities has very small effect on option price. But from the table 4.1 below we can see that, correlation among volatilities now are much more influential than before, as it now has a spill over effect on the forward rates, which comes from the correlation between the two innovations to forward rates and their volatilities. This makes sense intuitively, as highly correlated shocks to the volatilities (of different terms) will be accompanied by more correlated shocks to the forward rates and thus higher option price.

Correlation Between Forward Rate and its Volatility ρ

It has been noticed in the literature that, correlation between forward rate and its

Table 4.1: Price differences from different correlation parameter k_1 and k_2, where $c_i(u,v) = e^{-k_i|u-v|}$, $i = 1,\ 2$. Here, P_{Approx} is the price from the approximated distribution for the summary statistics, with calculated first two moments from section 4.2; P_{MC} is the price from the brute force Monte Carlo samples of the summary statistics; P^*_{Approx} is the price from approximated distribution for the summary statistics, with the two moments matching their MC samples; $TotalErr$ is the ratio of $RMSE$ and P_{MC}, where $RMSE$ is the square-root of average squared difference between P_{Approx} and P_{MC}; $AppoxErr$ is the ratio between $P^*_{Approx} - P_{MC}$ and P_{MC}. Other parameters values are chosen to be: $\alpha = -2, \beta = 1, \sigma_h = 1$ and $\rho = 0.5$; and the price of a \$10,000 notional at-the-money cap with $t = 1$ and $T = 1.25$ is considered and priced; procedures described in section 4.3.1 have been used to draw samples using Monte Carlo method, with sample size of 50,000; 200 grid points have been using between 0 and t, and between t and T as well.

k_1	k_2	P_{Approx}	P_{MC}	P^*_{Approx}	TotalErr	AppoxErr
-8	-1	3.52	3.49	3.55	1.1%	1.7%
-4	-1	3.84	3.81	3.87	1.0%	1.8%
-2	-1	4.04	4.02	4.07	0.9%	1.5%
-1	-1	4.14	4.14	4.20	0.7%	1.6%
-0.5	-1	4.22	4.20	4.26	0.9%	1.5%
-0.25	-1	4.26	4.23	4.28	0.9%	1.3%
-0.125	-1	4.30	4.25	4.31	1.4%	1.5%
0	-1	4.29	4.27	4.32	0.8%	1.4%
-1	-8	3.92	3.92	3.94	0.6%	0.8%
-1	-4	4.03	4.03	4.07	0.7%	1.2%
-1	-2	4.11	4.10	4.14	0.8%	1.3%
-1	-1	4.14	4.14	4.20	0.7%	1.6%
-1	-0.5	4.19	4.16	4.22	1.0%	1.6%
-1	-0.25	4.20	4.17	4.23	1.0%	1.5%
-1	-0.125	4.22	4.18	4.24	1.4%	1.7%
-1	0	4.22	4.18	4.24	1.1%	1.7%

volatility is necessary to produce a skewed implied volatility curve that has been frequently observed. To investigate if the same phenomenon exists in our more general framework, we will compute option prices and thus implied volatility curve under different correlation settings. From the figure 4.1 below, it can be seen that we are very flexible in terms of the shape of curve we can produce: skewed, smile shaped or reverse-skewed implied volatility curve can all be generated by tuning the correlation parameter. The lower correlation corresponds to skewed curve, the higher correlation corresponds to reverse-skewed curve, while something in the middle corresponds to smile shaped implied volatility curve.

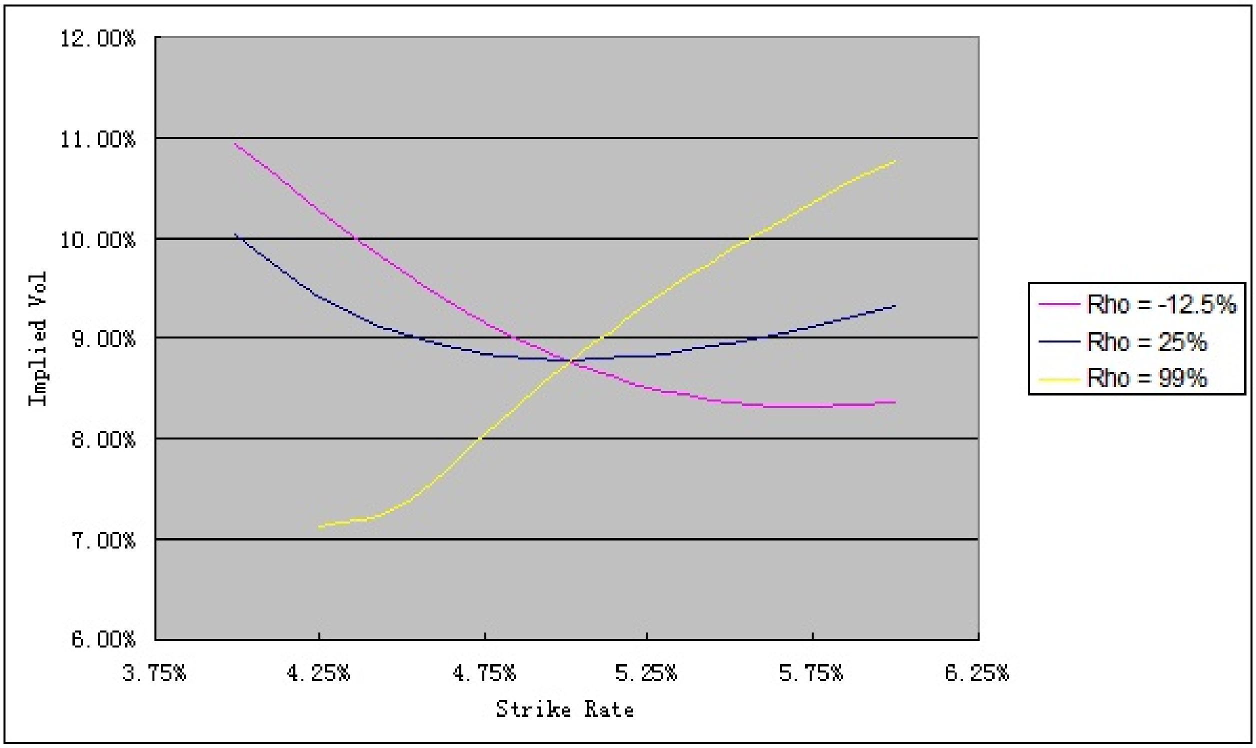

Figure 4.1: Implied Volatility for different $\rho's$. For different values of ρ, the pricing formula can produce different shape of Implied Volatility curve. For negative or small ρ, it will be skewed which is consistent with the observed price under normal market condition. And as ρ becomes larger, it will illustrate smile shape; and as ρ grows even larger, it will tilt further and can eventually become negatively skewed. This property of being able to produce different shapes of implied volatility has illustrated its flexibility in matching market prices and its potential in real world application.

Chapter 5

Model Calibration

5.1 Introduction

In this chapter, we will discuss the issue of calibrating random field forward rate model with stochastic volatilities(SV). There are generally three different approaches of calibrating the models, or in other words estimating the model parameters. The first option is the Least Square method. This method however, will usually require an explicit formula for assets (bond, cap or swaption), which generally is only available for Gaussian models. E.g., Pang(1998) calibrated Gaussian random field and Gaussian n-factor HJM model. Since there is a closed form solution for cap and semi-closed form for swaption, he used least square method from cap and swaption data. The second option is estimation by maximizing likelihood. Likelihood method is applicable to the stochastic volatility case, where Kalman filter can be used to maximize the quasi-likelihood function. E.g., Trolle and Schwartz(2009) used quasi-likelihood and extended Kalman filter to calibrate a very general HJM affine factor SV model with correlation between forward rate factors and volatility factors. Similarly, Han(2007) also used Kalman filter for zero-coupon bond prices to estimate a factor model with square-root process for the variance-covariance matrix without separating the correlation and the volatility specification. The third method of estimation will be using Bayesian MCMC method to obtain the posterior distribution for the parameters. E.g., Eraker (2004), Jacquier et al (1994) and Cheng et al (2008) used it for the estimation in the equity case; And Bester (2004) used it to estimate the Affine Factor and Affine Random Field Model for forward rates. Comparing MCMC method with the Kalman filter method, aside from one being Bayesian and the other one frequentist, they both are Monte Carlo based method. Kalman Filter is essentially still MLE. It will simulate a large amount of paths for the state variables, approximate the large integrals, and then solve for MLE. However since the estimation problem involves a large number of latent variables, and thus usually is very difficult to estimate, we can use Bayesian method to incorporate prior information to help the calibration. Especially from the literature of estimating OU process, there might be large bias, especially for the mean reverting speed parameter, in the estimation of parameters.

We will use Markov Chain Monte Carlo(MCMC) method and a similar data input as Bester (2004) to estimate a log-linear SV model. A number of authors have used the MCMC method to calibrate their models from real data, but few of them have actually carried out simulation studies to test the accuracy and performance of such method, which is the purpose of this chapter. We will examine in details a one factor SV model using simulated forward rate data to test the performance of the MCMC based method. First we describe the model we are going to calibrate and the chosen parameters for it. 100 different data sets each with a fixed sample size will be generated from the assumed model as input to the calibration step. OpenBUGS will then be used to implement the inference step where we will pick the sampling procedure for each of the parameters based on the forms of the posterior distributions. Finally, we test convergence of the

MCMC procedure and report results from the simulation study.

5.2 Simulation Study: One-factor SV model

In this section, we will try to calibrate one of the random field models with a one-factor stochastic volatility process, and use simulation technique to test the effectiveness of the calibration procedure. In essence, assuming that forward rate dynamics follow exactly the assumed model with known parameters, we can generate forward rate paths from the model for a certain period and then we will apply the calibration procedure and get the estimates for the model parameters. The performance of the calibration procedure can be tested by looking at the difference between the estimated parameters and their true values. Since bias can be introduced from a limited number of observations, we can either generate a very long series of data or we can fix a limited number of observations and repeat the data generation and calibration process for a large number of times. We will opt for the second option here because in reality only a limited number of samples are available, and we shall test the performance of the finite sample inference.

Consider the discretized version of the model with independent noise between forward rate and its volatility:

$$f_{t+\Delta}(T) - f_t(T) = \mu_t(T)\Delta + e^{\frac{h_t}{2}}(W_{t+\Delta}(T) - W_t(T)) \tag{5.1}$$

$$h_{t+\Delta} - h_t = (\alpha - \beta h_t)\Delta + \sigma_h(B_{t+\Delta} - B_t) \tag{5.2}$$

Here we approximate the forward rate curve $F_t = (f_t(T_1), \ldots, f_t(T_{N_T}))^T$ as a vector of length N_T, and μ_t will depend on which measure our model is built upon. I.e., under martingale measure it is a deterministic function of h_t and field correlation $c(u,v)$

following equation (2.3), and under physical measure it is not restricted. Here we assume we are estimating directly from observed forward rate curve and thus under physical measure. Also since our main focus will be on the volatility process and option price, we will assume that the curve has been de-meaned and thus $\mu(t)$ is always zero.

For simplicity here we will take the parametric form for $c(,)$ as

$$c(u, v) = e^{-k|u-v|} \tag{5.3}$$

To reduce the correlation between parameters and also notice that β needs to be constrained to make the discretized OU (AR(1)) process convergent, we will re parameterize equation (5.2) by introducing $\mu_h = \frac{\alpha}{\beta}$ and $\varphi^1 = 1 - \Delta\beta$ and thus equation (5.2) will become:

$$h_{t+\Delta} - \mu_h = \varphi^1(h_t - \mu_h) + \sigma_h(B_{t+\Delta} - B_t) \tag{5.4}$$

Since the convergence of (5.5) would require that $|\varphi^1| < 1$, we further introduce $\varphi = (\varphi^1 + 1)/2$ and impose the constrain that φ to be between 0 and 1, and equation (5.5) becomes:

$$h_{t+\Delta} = \mu_h + (2\varphi - 1)(h_t - \mu_h) + \sigma_h(B_{t+\Delta} - B_t) \tag{5.5}$$

We will use Bayesian method to calculate the distribution for the model parameters. I.e. given observed forward rate data and prior distributions on the parameters, we would like to generate posterior distributions for the parameters. The difficulty of the process lies in the un-observable volatility state variables $\{h_t\}$. We will treat them as latent variables and simulate samples from their conditional distributions along with other regular variables. In other words, what we need is $p(\mu_h, \varphi, \sigma_h, k|\{F_t\})$, but we will simulate from $p(\mu_h, \varphi, \sigma_h, k, \{h_t\} \mid \{F_t\})$ instead, and then the partial vector will be from the marginal distribution $p(\mu_h, \varphi, \sigma_h, k|\{F_t\})$ which is what we aim for. Markov Chain Monte Carlo (MCMC) method can be used to simulate from the distribution of this long vector. The full likelihood function for all observations is:

$$\begin{aligned} &p_0(\mu_h, \varphi, \sigma_h, k)p_0(h_0)p(\{h_t\}|\mu_h, \varphi, \sigma_h)p(\{F_t\}|\{h_t\}, k) \\ &= p_0(\mu_h)p_0(\varphi)p_0(\sigma_h)p_0(k)p_0(h_0)\prod_{j=1}^{N} p(h_{t_j}|h_{t_j-\Delta}, \mu_h, \varphi, \sigma_h)\prod_{j=1}^{N} p(F_{t_j}|h_{t_j-\Delta}, k) \end{aligned} \tag{5.6}$$

From the result of Gibbs sampling, in order to sample from the distribution of $p(\mu_h, \varphi, \sigma_h, k, \{h_t\} \mid \{F_t\})$, we need only sample sequentially as follows:

1. Set initial values for each of the parameters $\mu_h^{(0)}, \varphi^{(0)}, \sigma_h^{(0)}, k^{(0)}, h_0^{(0)}$.
2. draw a random sample of $k^{(i+1)}$ from $k|\{F_t\}, \mu_h^{(i)}, \varphi^{(i)}, \sigma_h^{(i)}, \{h_t^{(i)}\} \propto p_0(k)p(\{F_t\}|k, \{h_t^{(i)}\})$
3. draw random samples for $\sigma_h^{(i+1)}$ from $\sigma_h|\mu_h^{(i)}, \varphi^{(i)}, \{h_t^{(i)}\} \propto p_0(\sigma_h)p(\{h_t\}|\mu_h^{(i)}, \varphi^{(i)}, \sigma_h)$
4. draw random samples for $\mu_h^{(i+1)}$ from $\mu_h|\sigma_h^{(i)}, \varphi^{(i)}, \{h_t^{(i)}\} \propto p_0(\mu_h)p(\{h_t\}|\mu_h, \varphi^{(i)}, \sigma_h^{(i+1)})$
5. draw random samples for $\varphi^{(i+1)}$ from $\varphi|\sigma_h^{(i+1)}, \mu_h^{(i+1)}, \{h_t^{(i)}\} \propto p_0(\varphi)$ $p(\{h_t\}|\mu_h^{(i+1)}, \varphi, \sigma_h^{(i+1)})$
6. for each $t = t_1, \ldots, t_N$, draw a random sample of $h_t^{(i+1)}$ from $h_t|F_{t+\Delta}, h_{t-\Delta}^{(i+1)}, h_{t+\Delta}^{(i)}, \mu_h^{(i+1)}, \varphi^{(i+1)}, \sigma_h^{(i+1)}, k^{(i+1)}$ which has a density function proportional to
   $p(F_{t+\Delta}|h_t, k^{(i+1)})p(h_t|h_{t-\Delta}^{(i+1)}, \mu_h^{(i+1)}, \varphi^{(i+1)}, \sigma_h^{(i+1)})p(h_{t+\Delta}^{(i)}|h_t, \mu_h^{(i+1)}, \varphi^{(i+1)}, \sigma_h^{(i+1)})$

Then as i large enough, the sampled vector $(\mu_h^{(i)}, \varphi^{(i)}, \sigma_h^{(i)}, k^{(i)})$ will approximately come from the posterior distribution.

Now let's move on to the details of drawing samples from each of the steps. We will implement the MCMC inference using the OpenBUGS version of the BUGS (Bayesian inference Using Gibbs Sampling) software, within which we will specify the model and the sampling methods used in each step. The following subsections will describe the sampling procedure to be used in the calibration process. Note that, for the ease of read we will drop the iteration index for the parameters we are not updating when there are no confusions.

5.2.1 Sampling k

To sample from the posterior distribution $k|\{F_t\},\{h_t^{(i)}\} \propto p_0(k)p(\{F_t\}|k,\{h_t^{(i)}\})$ where

$$\begin{aligned} p(\{F_t\}|k,\{h_t^{(i)}\}) &\propto \prod_{j=1}^{N} |\Sigma_{t_j}|^{-\frac{1}{2}} e^{-\frac{1}{2}(F_{t_j+\Delta}-F_{t_j})^T \Sigma_{t_j}^{-1}(F_{t_j+\Delta}-F_{t_j})} \\ &= |\Sigma_{t_i}|^{-\frac{N}{2}} e^{-\frac{1}{2}\sum_{j=1}^{N}(F_{t_j+\Delta}-F_{t_j})^T \Sigma_{t_j}^{-1}(F_{t_j+\Delta}-F_{t_j})} \\ &\propto e^{-NN_T h_t/2} |\Sigma|_1^{-\frac{N}{2}} e^{-\frac{1}{2\Delta e^{h_t}}\sum_{j=1}^{N}((F_{t_j+\Delta}-F_{t_j}))^T \Sigma_1^{-1}(F_{t_j+\Delta}-F_{t_j})} \end{aligned} \tag{5.7}$$

where

$$\begin{aligned} \Sigma_t = Var(F_{t+\Delta}-F_t) &= \Delta e^{h_t} \left(\; e^{-k|T_m-T_n|} \; \right)_{m,n} \\ &\equiv \Delta e^{h_t} \Sigma_1 \end{aligned} \tag{5.8}$$

and thus $|\Sigma_t| = (\Delta e^{h_t})^{N_T}|\Sigma_1|$ and $\Sigma_t^{-1} = \Delta^{-1}e^{-h_t}\Sigma_1^{-1}$.

Since the distribution is highly non-standard which involves the calculation of norm and inverse of a matrix, no conjugate prior is available, nor is the log-likelihood guaranteed to be concave so that no envelop can be easily built to dominate it and thus Adaptive Rejection Sampling is also not applicable. Random walk Metropolis algorithm (e.g., Gelman etal (1996), Haario, Saksman and Tamminen (1999))can be used to do the sampling, but a more recent development by Neal (2003) introduced a more efficient slice sampling which has been used as the generic sampling method in BUGS.

The idea of slice sampling came from the observation that, in order to sample a variable k which has a distribution proportional to $f(k)$, we can sample uniformly over the plate of $U = \{(k, y) : 0 \le y \le f(k)\}$ which is just below the curve defined by $f(k)$. I.e., the joint distribution for (k, y) will be $p(k, y) \propto 1$ for $0 \le y \le f(k)$ and 0 otherwise. The marginal distribution for k will be $p(k) \propto \int_0^{f(k)} dy \propto f(k)$ which is exactly what we are looking for.

If we introduce an auxiliary variable y and we want to sample from the uniform distribution of (k, y), we can again follow the idea of Gibbs Sampling and simulate them one after another. Here is one proposal from Neal (2003) that can generate $(k^{(i+1)}, y^{(i+1)})$ from $(k^{(i)}, y^{(i)})$.

- draw a value $y^{(i+1)}$ uniformly from $(0, f(k^{(i)}))$, thereby introducing a horizontal slice $S = \{k : y^{(i+1)} \le f(k)\}$.

- Find an interval $I = (L,\ R)$ around $k^{(i)}$ that contains at least a large portion of the slice. For this step we choose to use the stepping out procedure proposed by Neal (2003). We define the interval by randomly positioning an interval of length w (chosen during the adaptive phrase) around $k^{(i)}$, and then expanding the interval in steps of w on both ends separately until they are outside of the slice.

- Draw a new point uniformly from within I until a point inside the slice S is found, and accept it as $k^{(i+1)}$. Points outside of the slice are used to shrink the interval.

5.2.2 Sampling μ_h, φ and σ_h

Since

$$\begin{aligned}
&\prod_{j=1}^{N} p(h_{t_j}|h_{t_j-\Delta}, \mu_h, \varphi, \sigma_h) \\
&= (2\pi)^{-\frac{N}{2}} \sigma_h^{-N} exp(\sum_{j=0}^{N-1} (h_{t_j+\Delta} - (2\varphi-1)h_{t_j} - 2(1-\varphi)\mu_h)^2/(2\sigma_h^2)) \qquad (5.9)
\end{aligned}$$

Given h_t, $t = t_1, \dots, t_n$, σ_h and φ, and assume a Normal prior as $N(\mu_0, \sigma_0^2)$ for μ_h, we keep the terms containing μ_h in equation (5.6) to get a posterior distribution of:

$$\begin{aligned}
p(\mu_h|\{h_t\}, \varphi, \sigma_h) &\propto p_0(\mu_h) \prod_{j=1}^{N} p(h_{t_j}|h_{t_j-\Delta}, \mu_h, \varphi, \sigma_h) \\
&\propto exp(-\frac{(\mu_h-\mu_0)^2}{2\sigma_0^2}) exp(-\frac{\sum_{j=0}^{N-1} (h_{t_j+\Delta} - (2\varphi-1)h_{t_j} - 2(1-\varphi)\mu_h)^2}{2\sigma_h^2}) \\
&\propto exp(-\frac{1}{2}[(\frac{1}{\sigma_0^2} + \frac{4N(1-\varphi)^2}{\sigma_h^2})\mu_h^2 - 2\mu_h(\frac{\mu_0}{\sigma_0^2} + \frac{2(1-\varphi)\sum h_{t_j+\Delta} - (2\varphi-1)h_{t_j}}{\sigma_h^2})]) \\
&\propto N(\mu_N, \sigma_N^2) \qquad (5.10)
\end{aligned}$$

here $\sigma_N^2 = (\frac{1}{\sigma_0^2} + \frac{4N(1-\varphi)^2}{\sigma_h^2})^{-1}$ and $\mu_N = \sigma_N^2(\frac{\mu_0}{\sigma_0^2} + \frac{2(1-\varphi)\sum h_{t_j+\Delta} - (2\varphi-1)h_{t_j}}{\sigma_h^2})$.

Similarly, denote $\tau = \frac{1}{\sigma_h^2}$ and assume a gamma prior $\tau \sim Gamma(g_0, \lambda_0)$, we can have a posterior for τ following Gamma distribution as

$$
\begin{aligned}
p(\tau|\{h_t\}, \mu_h, \varphi) &\propto p_0(\tau) \prod_{j=1}^{N} p(h_{t_j}|h_{t_j-\Delta}, \mu_h, \varphi, \sigma_h) \\
&\propto \tau^{g_0-1} e^{-\tau/\lambda_0} \tau^{N/2} exp(-\frac{\tau}{2} \sum_{j=0}^{N-1} (h_{t_j+\Delta} - (2\varphi - 1)h_{t_j} - 2(1-\varphi)\mu_h)^2) \\
&\propto Gamma(g_0 + N/2, (1/\lambda_0 + \frac{1}{2} \sum_{j=0}^{N-1} (h_{t_j+\Delta} - (2\varphi - 1)h_{t_j} - 2(1-\varphi)\mu_h)^2)^{-1})
\end{aligned}
\tag{5.11}
$$

Since we need the discretized AR(1) process to be convergent which is equivalent of φ being between 0 and 1, we will assume a beta prior for it as $Beta(b_0, b_1)$, and thus it has a posterior of

$$
\begin{aligned}
p(\varphi|\mu_h, \sigma_h, \{h_t\}) &\propto p_0(\varphi) \prod_{j=1}^{N} p(h_{t_j}|h_{t_j-\Delta}, \mu_h, \varphi, \sigma_h) \\
&\propto \varphi^{b_0-1}(1-\varphi)^{b_1-1} exp(-\frac{1}{2\sigma_h^2} \sum_{j=0}^{N-1} (h_{t_j+\Delta} - (2\varphi - 1)h_{t_j} - 2(1-\varphi)\mu_h)^2) \\
&\propto \varphi^{b_0-1}(1-\varphi)^{b_1-1} exp(-\frac{2\varphi^2}{\sigma_h^2} \sum_{j=0}^{N-1} (h_{t_j} - \mu_h)^2 + \frac{2\varphi}{\sigma_h^2} \sum_{j=0}^{N-1} (h_{t_j} - \mu_h)(h_{t_j+\Delta} + h_{t_j} - 2\mu_h)) \\
&\propto \varphi^{b_0-1}(1-\varphi)^{b_1-1} exp(-A_N\varphi^2 + B_N\varphi)
\end{aligned}
\tag{5.12}
$$

Here $A_N = -\frac{2}{\sigma_h^2} \sum_{j=0}^{N-1} (h_{t_j} - \mu_h)^2$ and $B_N = \frac{2}{\sigma_h^2} \sum_{j=0}^{N-1} (h_{t_j} - \mu_h)(h_{t_j+\Delta} + h_{t_j} - 2\mu_h)$. Since the distribution is non-standard, slice sampling similar to the sampling of k can be used to simulate from this distribution.

Note that, it is known in the literature that estimation of the parameter β is very difficult, and there is usually an upward bias when using MLE or LS. We also expect the same bias from Bayesian method. To minimize the effect of this effect, we may need to apply a somewhat informative prior on the parameters. This prior will depend on experience and the type of asset to calibrate.

5.2.3 Sampling for $h_t,\ t = t_1, \dots, t_T$

In order to sample a series of $\{h_t\}$, we can follow the concept of Gibbs sampling by simulating them one by one. Assume we have updated until $\{h_{t-\Delta}\}$, then to sample h_t, we have its posterior distribution as $p(h_t|F_t, h_{t-\Delta}^{(i+1)}, h_{t+\Delta}^{(i)}, \mu_h, \varphi, \sigma_h, k) \propto p(F_t|h_t, k)p(h_t|h_{t-\Delta}^{(i+1)}, \mu_h, \varphi, \sigma_h)p(h_{t+\Delta}^{(i)}|h_t, \mu_h, \varphi, \sigma_h)$, where

$$
\begin{aligned}
p(F_t|h_t, k) &\propto |\Sigma_t|^{-\frac{1}{2}} e^{-\frac{1}{2}(F_{t+\Delta}-F_t)^T \Sigma_t^{-1}(F_{t+\Delta}-F_t)} \\
&\propto e^{-N_T h_t/2} e^{-\frac{1}{2\Delta e^{h_t}}(F_{t+\Delta}-F_t)^T \Sigma_1^{-1}(F_{t+\Delta}-F_t)} \\
p(h_t|h_{t-\Delta}^{(i+1)}, \mu_h, \varphi, \sigma_h) &\propto e^{-\frac{1}{2\Delta\sigma_h^2}(h_t-(2\varphi-1)h_{t-\Delta}-2(1-\varphi)\mu_h)^2} \\
p(h_{t+\Delta}^{(i)}|h_t, \mu_h, \varphi, \sigma_h) &\propto e^{-\frac{1}{2\Delta\sigma_h^2}(h_{t+\Delta}-(2\varphi-1)h_t-2(1-\varphi)\mu_h)^2}
\end{aligned}
$$

Since the posterior distribution is again non-standard, slice sampling can be used to sample from it. The procedure will be the same as the sampling of k.

5.3 Simulation Results

5.3.1 Set up

Detailed simulation process for the model setting of (5.1) - (5.3) will be the following:

- Set model parameters as $k = 1$, $\alpha = -2$, $\beta = 1$, $\sigma_h = 1$ and thus $\mu_h = \frac{\alpha}{\beta} = -2$. Also assume that the forward rate curve is observable at the fix terms (0, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 8, 10).

- Generate a path for h_t and then F_t with 500 observations for 10 years, and thus

$\Delta = 0.02$, $T = 10$ and $\varphi = 0.99$;

- Calibrate the model from values of F_t alone with prior distribution for the parameters chosen as $k \sim Gamma(0.1,\ 0.1)$, $\mu_h \sim Norm(-1,\ 20)$, $\varphi \sim Beta(20,\ 1.5)$ and $\sigma_h^2 \sim IG(0.1,\ 0.1)$.

- Repeat the last three steps for 100 times.

5.3.2 Convergence Diagnostic

To test if the MCMC procedure is convergent or not, we look at one of the estimation results in details:

- Three different sets of initial values have been used. It can be seen from Figure 5.1 that after the burn in period, the three different chains converge and they mix well with each other. The initial values does not seem to affect the posterior distribution;

- Autocorrelation: from Figure 5.2 there are still some autocorrelations for k after the thin of 10, autocorrelations for other parameters are much better.

- BGR plot of Figure 5.3: it can be seen that all three lines stabilize after the burn in period, and the red line converges to 1, which is a good indication of convergence based on Brooks and Gelman (1998)

- running mean and quantile in Figure 5.4: they are all stable after the burn in period.

The above observations all indicate that the MCMC samples have converged reasonably well.

5.3.3 Estimation Result

Table 5.1 and Figure 5.5 summarize the results from the 100 reps of the calibration problem. Here "True" column gives the actual parameter values we choose for the study; "Avg(Mean)" column is the average of the 100 posterior mean estimates for each replication, the distance between this and the True column measures the magnitude of the bias; "Std(Mean)" is the standard error of the 100 posterior mean estimates and it indicates how far the posterior mean could be from the actual mean in each replication; "Avg(SD)" is the average of the 100 posterior standard deviation estimates, which is just the average of posterior standard deviation estimates.

It can be seen that:

- The parameter on the forward rate correlation and the mean level of the volatility factor are very well estimated, with little bias and small deviate from the true value in each replication.

- The volatility of the volatility is hard to estimate with high standard deviation, but it is reasonably unbiased.

- The mean reverting speed parameter is very difficult to estimate with big bias and big variance, which is consistent with the O-U process estimation in the literature (e.g., Tang and Chen (2009), Yu (2009)).

In conclusion, MCMC method provides a good way of estimating stochastic volatility model with hidden state variables $\{h_t\}$. The estimates on model parameters are mostly unbiased and accurate except that it provides no magic for the estimation of mean reversion speed parameter for the OU process, and will still be somewhat biased and skew to the right.

	True	Avg(Mean)	Std(Mean)	Avg(SD)	Coverage
k	0.1	0.10	0.005	0.005	98%
μ_h	-2.0	-2.03	0.19	0.21	94%
β	1.0	1.45	0.67	0.62	63%
σ_h	1.0	1.04	0.13	0.13	93%

Table 5.1: Summary statistics from the 100 reps. Estimation on the correlation parameter among the forward rates (k) and the mean volatility level (μ_h) are unbiased with small error, which is reasonable as we have direct observation on the forward rates; The estimation on β and σ_h will be more difficult as they are not directly observable, especially estimate for *beta* has a large bias and large standard error, which is consistent with the literature that mean reverting parameter of a OU process is very difficult to estimate.

5.4 Appendix: BUGS code

$$\left.\begin{array}{l}\left.\left.\text{yisigma}_{2,t,i,j} = ((\text{isig}_{2,i,j}/e^{\theta_t})/\text{sqrtdt})/\text{sqrtdt}\right\}1 \le j \le \text{NT}\right\}1 \le i \le \text{NT} \\ y_{t,1\ldots\text{NT}} \sim \text{dmnorm}(\text{my}_{1\ldots\text{NT}}, \text{yisigma}_{2,t,1\ldots\text{NT},1\ldots\text{NT}})\end{array}\right\}1 \le t \le N$$

$$\begin{aligned}
\mu &\sim \text{dnorm}(-1, 0.05) \\
\text{phi}^\star &\sim \text{dbeta}(20, 1.5) \\
\text{ivar} &\sim \text{dgamma}(0.1, 0.1) \\
\text{itau}_2 &= (\text{ivar}/\text{sqrtdt})/\text{sqrtdt} \\
\phi &= 2 \cdot \text{phi}^\star - 1 \\
\beta &= ((1-\phi)/\text{sqrtdt})/\text{sqrtdt} \\
\text{sig}_h &= \sqrt{1/\text{ivar}} \\
\theta_0 &\sim \text{dnorm}(\mu, \text{itau}_2) \\
\text{thmean}_1 &= \mu + \phi \cdot (\theta_0 - \mu) \\
\theta_1 &\sim \text{dnorm}(\text{thmean}_1, \text{itau}_2) \\
k &\sim \text{dgamma}(0.1, 0.1)
\end{aligned}$$

$$\left.\left.\text{sig}_{2,i,j} = e^{(-1)\cdot k\cdot \text{abs}(\text{vt}_i - \text{vt}_j)}\right\}1 \le j \le \text{NT}\right\}1 \le i \le \text{NT}$$

$$\text{isig}_{2,1\ldots\text{NT},1\ldots\text{NT}} = \text{inverse}(\text{sig}_{2,1\ldots\text{NT},1\ldots\text{NT}})$$

$$\left.\begin{array}{lcl} \text{thmean}_t & = & \mu + \phi \cdot (\theta_{t-1} - \mu) \\ \theta_t & \sim & \text{dnorm}(\text{thmean}_t, \text{itau}_2) \end{array}\right\} 2 \leq t \leq N$$

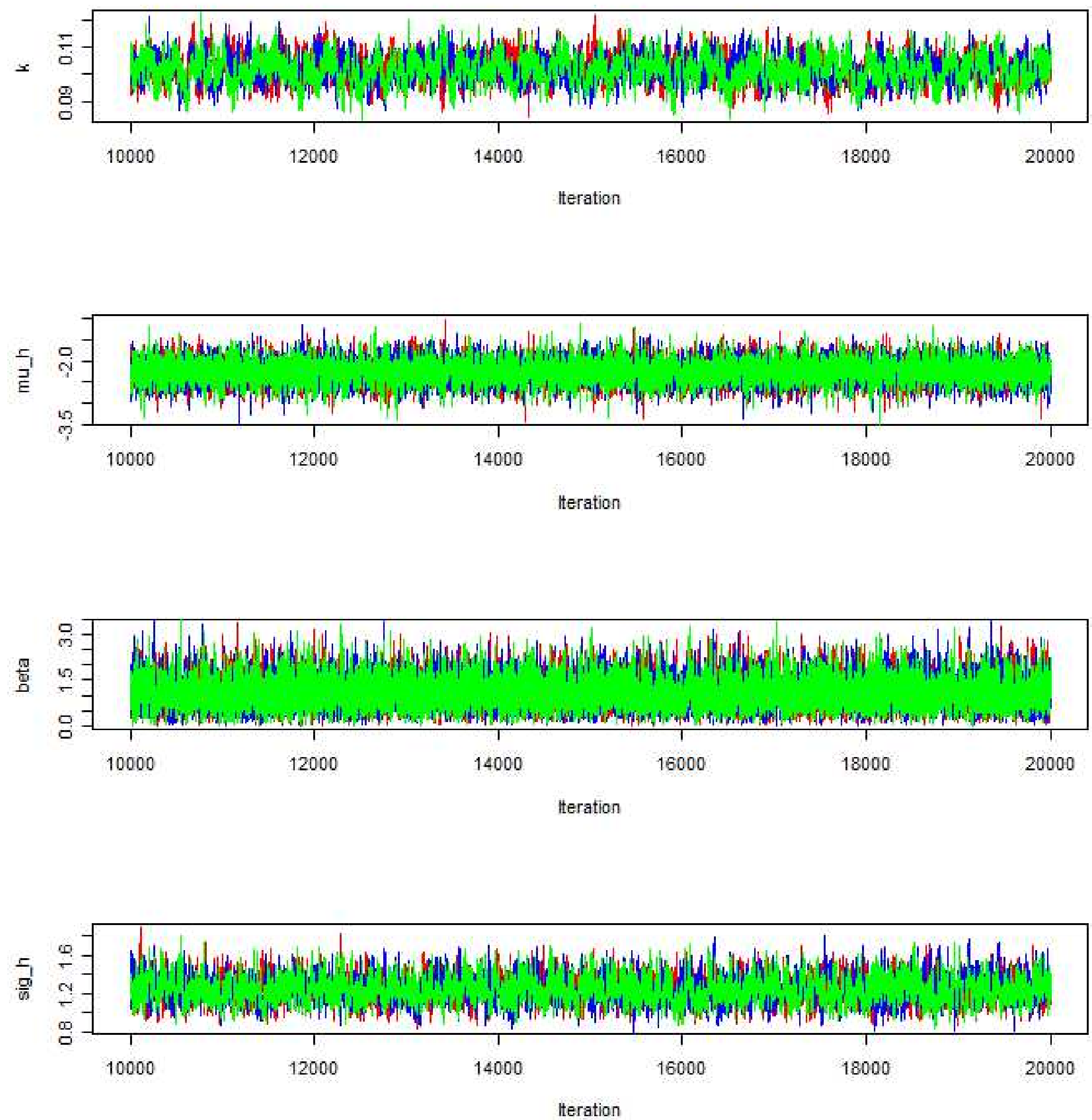

Figure 5.1: Trace plot for each of the parameters after discarding the first 10,000 samples, where each sample is retained only every 10th iteration, i.e. thinning of 10. Three chains have been produced here with different initial values. The three chains are very well mixed and it is a good indication of convergence for the MCMC procedure.

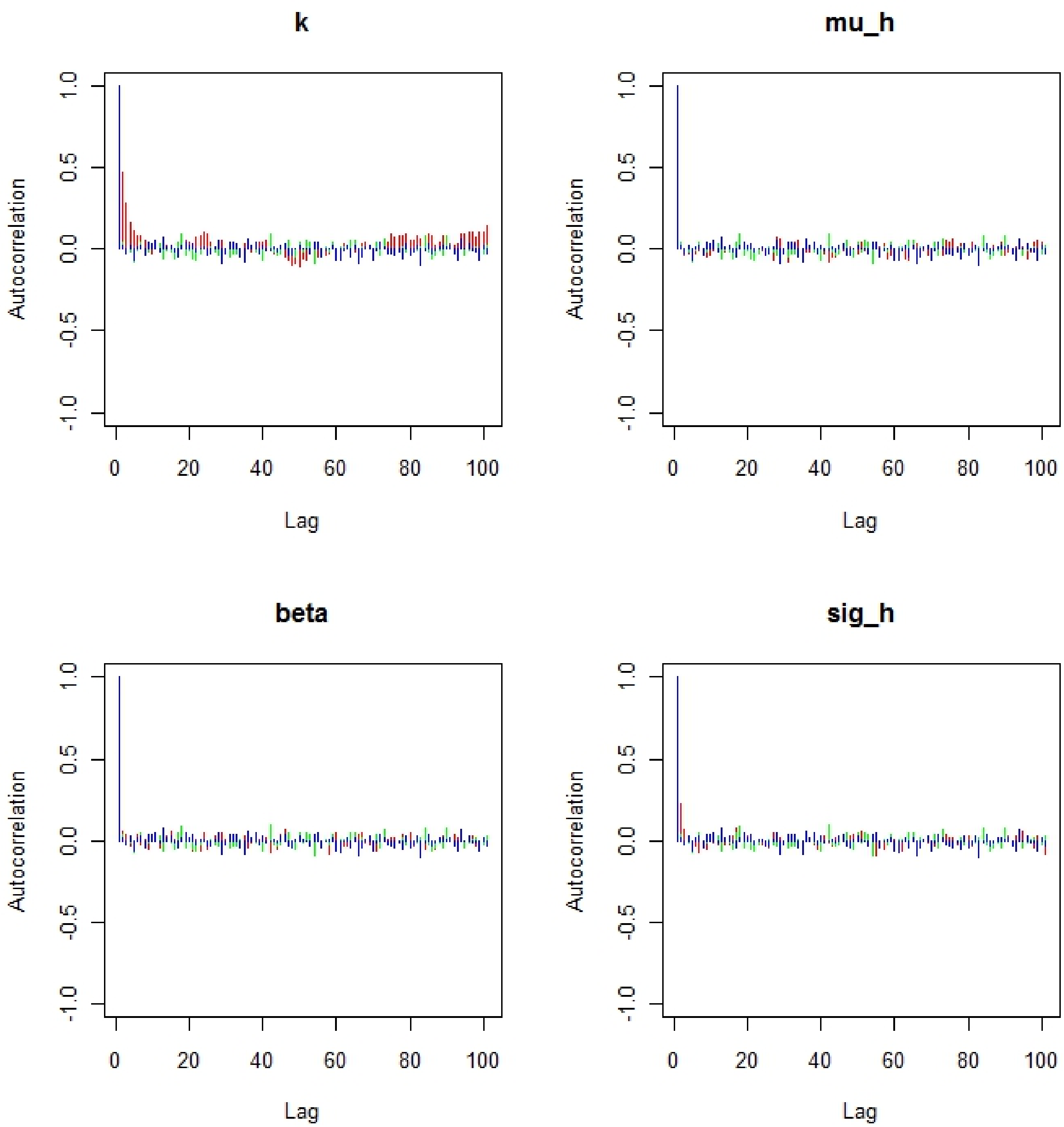

Figure 5.2: Auto-correlation between draws of various lags for a series of every 10th samples. Here the samples are obtained by keeping every 10th-iteration and also discard the first 10,000 samples as burn in. Even after this effectively thinning of 100, there is still some autocorrelation between the samples of k, while other parameters are much better. This high autocorrelation between MCMC draws will reduce the effective sample size, and thus longer chain is needed for a more accurate posterior distribution.

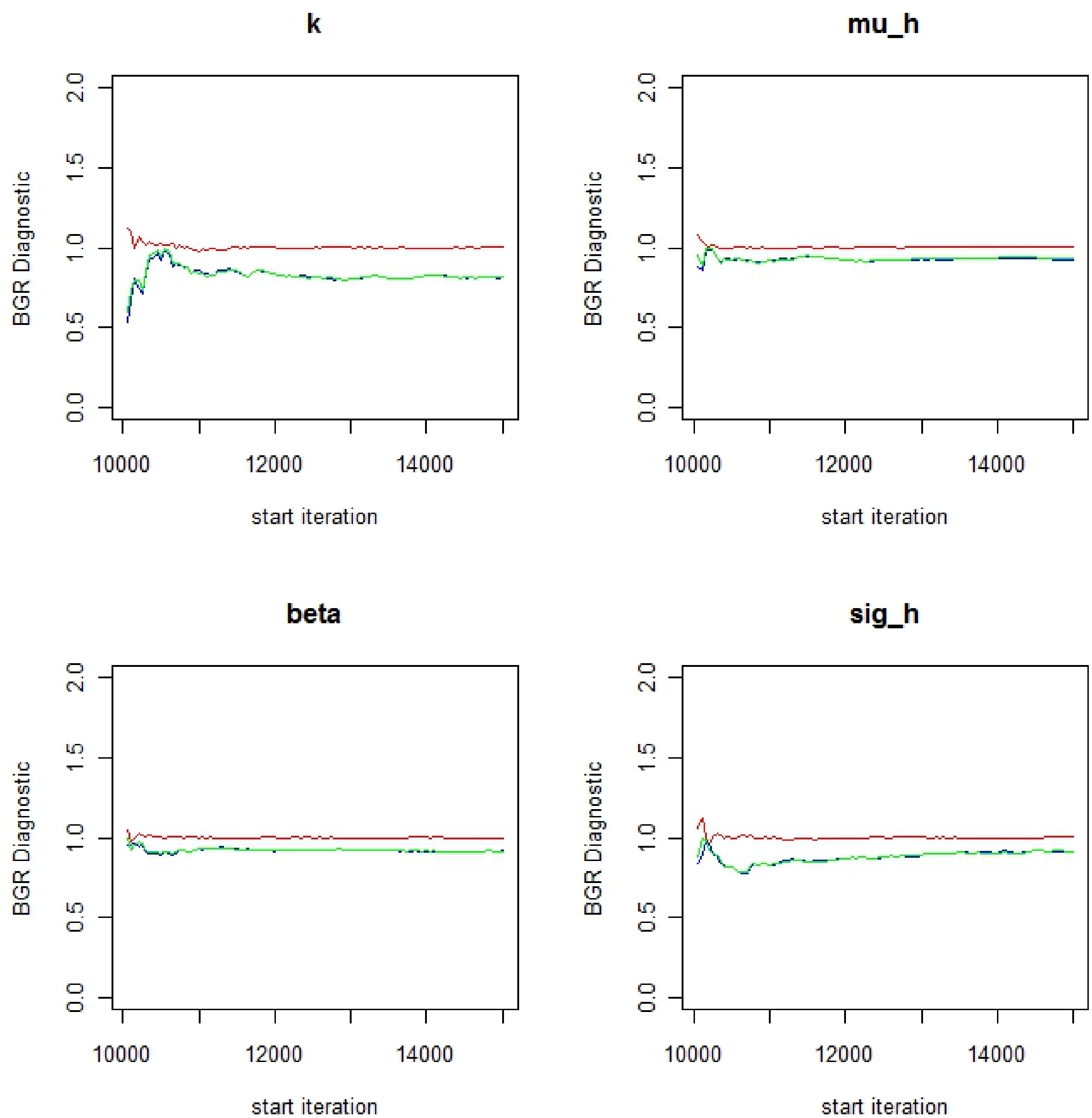

Figure 5.3: BGR plot: this comes from the idea of ANOVA where variance estimate from within chains should be close to the variance estimate from across chains. And thus a ratio of the two estimates (red line) closing to 1 is a good indication of convergence.

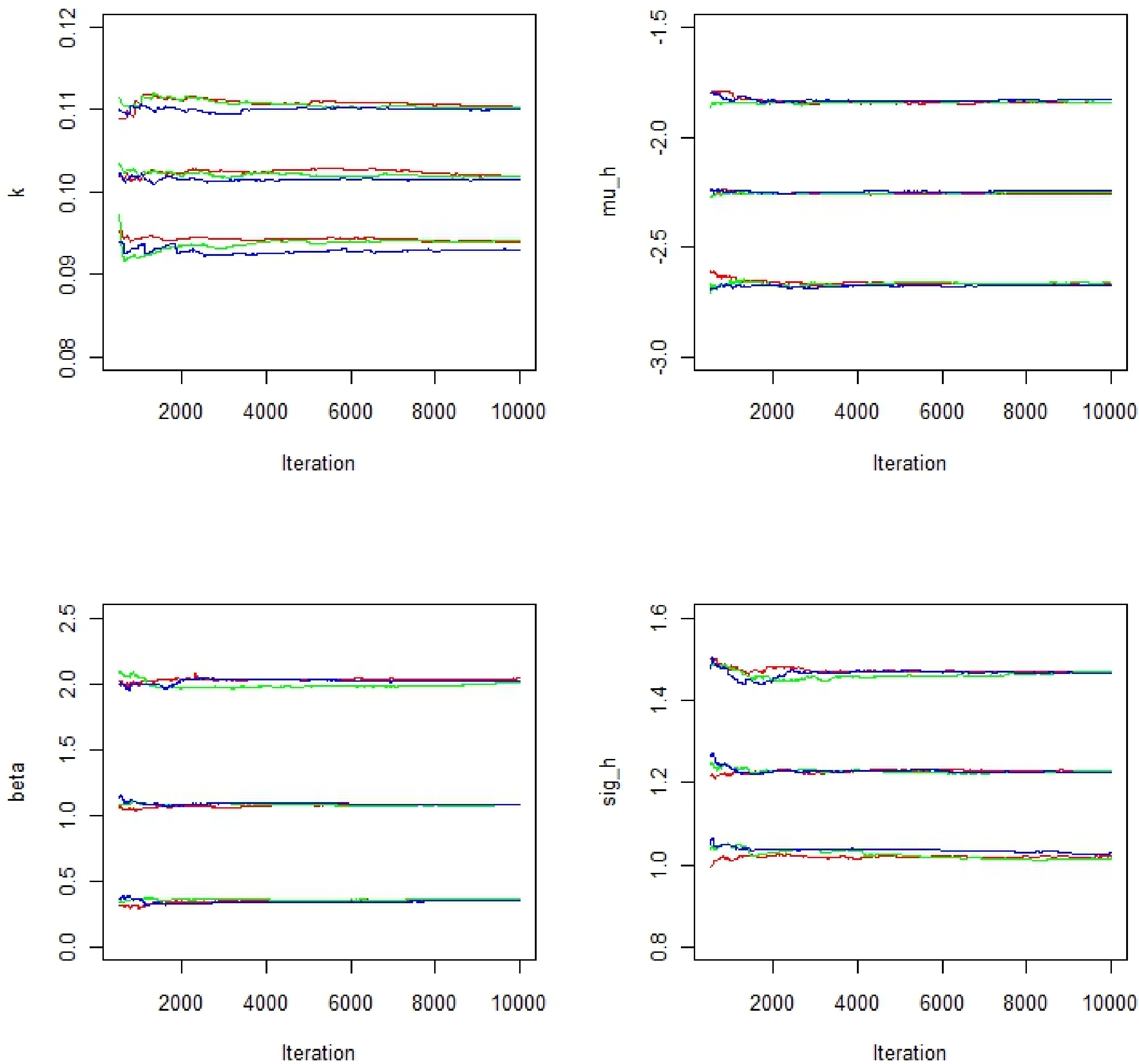

Figure 5.4: Running quantile plot: 5%, 50% and 95% percentile of the retained samples upto each point. Stable percentiles for each variable are also a good indication of convergence.

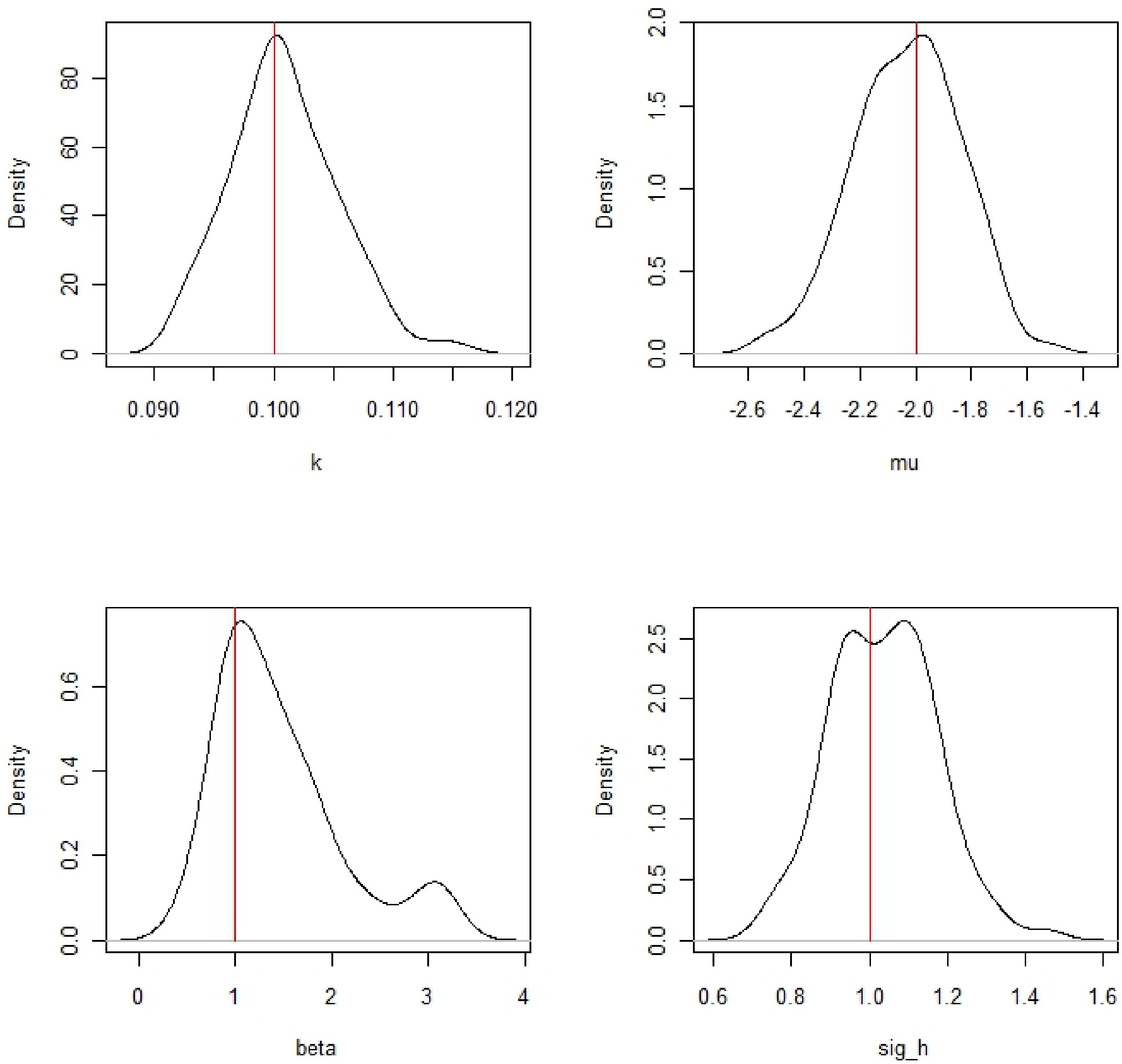

Figure 5.5: Density plots for the posterior mean estimates from the 100 replications. Here the red lines mark the true value for each of the parameters. A density more concentrated around the true value indicates the MCMC procedure described here in general provides good estimate for the parameter. It can be seen that estimation for k is both unbiased and accurate, estimates for μ_h and σ_h are unbiased in general, but they can still be pretty far from the true value and thus a larger number of observation might be needed to improve the estimation; While estimate for $beta$ is somewhat biased with skew to the right.

Chapter 6

Future Work

In the preceding chapters, we mainly discuss the issue of cap pricing under random field interest rate model with random field volatility structure and have shown that the model is very flexible to match the various features exhibited in the cap market. To make the model complete and comprehensive, more work need to be done so that it can price other interest rate derivatives (e.g. swaption) quickly, so that the model can be calibrated using both bond prices and option prices, so that, ideally, the model can be used to predict future interest rates movements. There are still a lot of open questions need to be addressed and we comment here about some potential future work:

- Swaption pricing: the success of our approximation method for cap price has relied on its closed form solution under deterministic volatility. Unfortunately there does not exist a closed form solution but rather an approximated formula for swaption price, which was proved to be accurate under Gaussian model. A natural next step will be to extend our method into swaption price utilizing the approximated formula. Its accuracy remains to be seen as there will be two steps of approximating now which could amplify error on the way.

- Incomplete market: following discussion in the Market Completeness section of Chapter 2, it is obvious that the model of our discussion results in an incomplete market which in turn implies that there may not be a unique martingale measure. One issue that needs to be addressed is to find a market price of risk process for the two random field sources that can hopefully preserve the structure of the model after changing of measure. There is a "extended affine" market price of risk specification suggested by Cheredito, Filipovic and Kimmel (2007) that will preserve the affine structure of a factor model, however, no such specification exists for random field model as far as we are aware. It will be very interesting to see if there exists such a structure preserving market price of risk process in the random field case.

- The most general random field model will of course contain the factor HJM model as a special case. However, for a implementable random field model, we have to make some simplifying assumptions, e.g. assuming some parametric form for the correlation structure, which may not necessarily be more flexible than their HJM counterparts. So another interesting extension is to find a correlation structure that is both easy to fit and flexible. This is a nontrivial task as not all functions satisfy the requirement of being one as discussed in Chapter 2. Best (2004) used nonparametric method to calibrate the correlation function which is obviously very flexible, however it is probably difficult to have a good estimate with the existence of stochastic volatility. Another interesting extension is to allow the correlation function to be random as well. This has been possible in the factor HJM model, e.g. in Han (2007) and Trolle and Schwartz (2009), where they start with random covariance and thus it is natural that the correlation can be random.

Bibliography

Ball, C.A. and W.N. Torous. 1999. "The Stochastic Volatility of Short-Term Interest Rates: Some International Evidence." *The Journal of Finance* 54(6):2339–2359.

Bensoussan, A. 1984. "On the theory of option pricing." *Acta applicandae mathematicae* 2(2):139–158.

Bester, C.A. 2004. Random Field and Affine Models for Interest Rates: An Empirical Comparison. Technical report Working Paper.

Björk, T., G. Di Masi, Y. Kabanov and W. Runggaldier. 1997. "Towards a general theory of bond markets." *Finance and Stochastics* 1(2):141–174.

Brigo, D. and F. Mercurio. 2006. *Interest rate models: theory and practice: with smile, inflation, and credit.* Springer Verlag.

Brooks, S.P. and A. Gelman. 1998. "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics* pp. 434–455.

Carmona, R. and M. Tehranchi. 2006. *Interest rate models: an infinite dimensional stochastic analysis perspective.* Springer Verlag.

Casassus, J., P. Collin-Dufresne and B. Goldstein. 2005. "Unspanned stochastic volatility and fixed income derivatives pricing." *Journal of Banking & Finance* 29(11):2723–2749.

Chen, R.R. and L. Scott. 2001. "Stochastic volatility and jumps in interest rates: An empirical analysis." *Working Paper* .

Cheng, A., A.R. Gallant, C. Ji and B.S. Lee. 2008. "A Gaussian approximation scheme for computation of option prices in stochastic volatility models." *Journal of Econometrics* 146(1):44–58.

Cheridito, P., D. Filipovic and R.L. Kimmel. 2007. "Market price of risk specifications for affine models: Theory and evidence." *Journal of Financial Economics* 83(1):123–170.

Chernov, M., A. Ronald Gallant, E. Ghysels and G. Tauchen. 2003. "Alternative models for stock price dynamics." *Journal of Econometrics* 116(1-2):225–257.

Chiarella, C. and O.K. Kwon. 2000. "A Class of Heath-Jarrow-Morton Term Structure Models with Stochastic Volatility." *Research Paper Series* .

Chib, S., F. Nardari and N. Shephard. 2002. "Markov chain Monte Carlo methods for stochastic volatility models." *Journal of Econometrics* 108(2):281–316.

Clark, P.K. 1973. "A subordinated stochastic process model with finite variance for speculative prices." *Econometrica* 41(1):135–155.

Collin-Dufresne, P. and R.S. Goldstein. 2003. "Generalizing the Affine Framework to HJM and Random Field Models." *Working Paper* .

Collin-Dufresne, P., R.S. Goldstein and C.S. Jones. 2009. "Can interest rate volatility be extracted from the cross section of bond yields?" *Journal of Financial Economics* 94(1):47–66.

Cox, J.C., J.E. Ingersoll Jr and S.A. Ross. 1985. "A theory of the term structure of interest rates." *Econometrica* 53(2):385–407.

Dai, Q. and K. Singleton. 2003. "Term structure dynamics in theory and reality." *Review of Financial Studies* 16(3):631.

Danielsson, J. 1994. "Stochastic volatility in asset prices. Estimation with simulated maximum likelihood." *Journal of Econometrics* 64(1-2):375–400.

Duffie, D., D. Filipović and W. Schachermayer. 2003. "Affine processes and applications in finance." *Annals of Applied Probability* pp. 984–1053.

Eraker, B. 2004. "Do stock prices and volatility jump? Reconciling evidence from spot and option prices." *Journal of Finance* pp. 1367–1403.

Gallant, A.R., D.A. Hsieh and G.E. Tauchen. 1991. On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-1983. In *Nonparametric and semiparametric methods in econometrics and statistics, Fifth International Symposium in Economic Theory and Econometrics.* pp. 199–240.

Gelfand, A.E. and A.F.M. Smith. 1990. "Sampling-based approaches to calculating marginal densities." *Journal of the American statistical association* pp. 398–409.

Gelman, A., G. Roberts and W. Gilks. 1996. "Efficient metropolis jumping hules." *Bayesian statistics* 5:599–608.

Ghysels, E., A.C. Harvey and E. Renault. 1996. "Stochastic volatility, chapter 14 in GS Maddala and CR Rao (eds), Handbook of Statistics, Vol. 14.".

Goldstein, RS. 2000. "The term structure of interest rates as a random field." *Review of Financial Studies* 13(2):365–384.

Haario, H., E. Saksman and J. Tamminen. 1999. "Adaptive proposal distribution for random walk Metropolis algorithm." *Computational Statistics* 14(3):375–396.

Han, B. 2007. "Stochastic volatilities and correlations of bond yields." *JOURNAL OF FINANCE-NEW YORK-* 62(3):1491.

Heath, D., R. Jarrow and A. Morton. 1992. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation." *Econometrica* 60(1):77–105.

Heidari, M. and L. Wu. 2003. "Are interest rate derivatives spanned by the term structure of interest rates?" *Journal of Fixed Income* .

Heston, S.L. 1993. "A closed-form solution for options with stochastic volatility with applications to bond and currency options." *Review of Financial Studies* pp. 327–343.

Hull, J. and A. White. 1987. "The pricing of options on assets with stochastic volatilities." *Journal of finance* pp. 281–300.

Hull, J. and A. White. 1990. "Pricing interest-rate-derivative securities." *Review of financial studies* 3(4):573.

Hull, J. and A. White. 1993. "Efficient procedures for valuing European and American path-dependent options." *The Journal of Derivatives* 1(1):21–31.

Jacquier, E., N.G. Polson and P.E. Rossi. 1994. "Bayesian analysis of stochastic volatility models." *Journal of Business & Economic Statistics* 12(4):371–389.

Jacquier, E., N.G. Polson and P.E. Rossi. 2004. "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors." *Journal of Econometrics* 122(1):185–212.

Jagannathan, R., A. Kaplin and S. Sun. 2003. "An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices." *Journal of Econometrics* 116(1-2):113–146.

James, J. and N. Webber. 2000. *Interest Rate Modelling.* John Wiley & Sons.

Jarrow, R., H. Li and F. Zhao. 2007. "Interest Rate Caps Smile Too! But Can the LIBOR Market Models Capture the Smile?" *The Journal of Finance* 62(1):345–382.

Kennedy, DP. 1994. "The Term Structure of Interest Rate as a Gaussian Random Field." *Mathematical Finance* 4(3):247–258.

Kennedy, DP. 1997. "Characterizing Gaussian Models of the Term Structure of Interest Rates." *Mathematical Finance* 7(2):107–118.

Kerkhof, J., A. Pelsser and Center for Economic Research (Katholieke Universiteit Brabant). 2002. "Observational equivalence of discrete string models and market models." *Journal of Derivatives* 10(1):55–61.

Kim, S., N. Shephard and S. Chib. 1998. "Stochastic volatility: likelihood inference and comparison with ARCH models." *The Review of Economic Studies* 65(3):361–393.

Kimmel, R.L. 2004. "Modeling the term structure of interest rates: A new approach." *Journal of Financial Economics* 72(1):143–183.

Krylov, N.V., M. Röbckner, J. Zabczyk and C.I. Matematico. 1999. *Stochastic PDE's and Kolmogorov equations in infinite dimensions*. Springer-Verlag.

Li, H. and F. Zhao. 2006. "Unspanned stochastic volatility: Evidence from hedging interest rate derivatives." *The Journal of Finance* 61(1):341–378.

Longstaff, F.A., P. Santa-Clara and E.S. Schwartz. 2001. "The relative valuation of caps and swaptions: Theory and empirical evidence." *The Journal of Finance* 56(6):2067–2109.

Lunn, D., D. Spiegelhalter, A. Thomas and N. Best. 2009. "The BUGS project: Evolution, critique and future directions." *Statistics in medicine* 28(25):3049–3067.

Neal, R.M. 2003. "Slice sampling." *The annals of statistics* 31(3):705–741.

Pan, J. 2002. "The jump-risk premia implicit in options: Evidence from an integrated time-series study." *Journal of Financial Economics* 63(1):3–50.

Pang, K. 1998. "Calibration of Gaussian heath, Jarrow and Morton and random field interest rate term structure models." *Review of Derivatives Research* 2(4):315–345.

Rebonato, R. 1999. "On the pricing implications of the joint log-normality assumption for the cap and swaption markets." *Journal of computational Finance* 2(3):30–52.

Renault, E. and N. Touzi. 1996. "Option Hedging and Implied Volatilities in a Stochastic Volatility Model." *Mathematical Finance* 6(3):279–302.

Ritchken, P. and L. Sankarasubramanian. 1995. "Volatility Structures Of Forward Rates And The Dynamics Of The Term Structure." *Mathematical Finance* 5(1):55–72.

Roberts, G.O. and J.S. Rosenthal. 2001. "Optimal scaling for various Metropolis-Hastings algorithms." *Statistical Science* 16(4):351–367.

Santa-Clara, P. and D. Sornette. 2001. "The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks." *The Review of Financial Studies* 14(1):149–185.

Scott, L.O. 1987. "Option pricing when the variance changes randomly: Theory, estimation, and an application." *Journal of Financial and Quantitative analysis* 22(4):419–438.

Scott, L.O. 1997. "Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods." *Mathematical Finance* 7(4):413–426.

Tang, C.Y. and S.X. Chen. 2009. "Parameter estimation and bias correction for diffusion processes." *Journal of Econometrics* 149(1):65–81.

Tauchen, G.E. and M. Pitts. 1983. "The price variability-volume relationship on speculative markets." *Econometrica: Journal of the Econometric Society* pp. 485–505.

Taylor, S.J. 1982. "Financial returns modelled by the product of two stochastic processesa study of the daily sugar prices 1961-75." *Time Series Analysis: Theory and Practice* 1:203–226.

Trolle, A.B. and E.S. Schwartz. 2009. "A general stochastic volatility model for the pricing of interest rate derivatives." *Review of Financial Studies* 22(5):2007–2057.

Vasicek, O. 1977. "An equilibrium characterization of the term structure." *Journal of financial economics* 5(2):177–188.

Yu, J. 2009. "Bias in the estimation of the mean reversion parameter in continuous time models.".

Zabczyk, J. 1999. "Parabolic equations on Hilbert spaces." *Stochastic PDEs and Kolmogorov Equations in Infinite Dimensions* pp. 117–239.